

ДЕРЖАВНА НАУКОВА УСТАНОВА  
«ІНСТИТУТ ІНФОРМАЦІЇ, БЕЗПЕКИ І ПРАВА  
НАЦІОНАЛЬНОЇ АКАДЕМІЇ ПРАВОВИХ НАУК  
УКРАЇНИ»

**Ланде Д.В., Фурашев В.М., Брайчевський С.М.**

**ІНФОРМАТИКА  
ПАРЛАМЕНТСЬКОГО КОНТРОЛЮ**

**ПОСІБНИК**

Київ – 2022

УДК 004.056:340.132.1+316.324.8

**Рекомендовано до друку**

Вченою радою Державної наукової установи «Інститут інформації, безпеки і права Національної академії правових наук України» (протокол від 14 грудня 2022 р. № 11)

**Рецензенти:**

*Довгань О.Д.* – доктор юридичних наук, професор, заслужений діяч науки і техніки України

*Новіков О.М.* – доктор технічних наук, професор, член-кореспондент НАН України

**Інформатика парламентського контролю: посібник /**

Ланде Д.В., Фурашев В.М., Брайчевський С.М. – Київ: ТОВ "Інжиніринг", 2022. – 256 с.

Видання присвячене розгляду актуальних теоретико-правових та прикладних проблем підвищення ефективності парламентського контролю, як одного факторів народовладдя, та шляхів їх вирішення.

Видання розраховане на фахівців, науково-педагогічних працівників, аспірантів, докторантів, студентів та курсантів закладів вищої освіти, представників державних органів та органів місцевого самоврядування, установ та організацій.

ISBN 978-966-2344-80-6

© Ланде Д.В., Фурашев В.М.,  
Брайчевський С.М., 2022

## **ЗМІСТ**

<b>ПЕРЕДМОВА</b>	5
<b>I. ПАРЛАМЕНТСЬКІЙ КОНТРОЛЬ</b>	7
1.1. Поняття «парламентській контроль» та нормативно-правова база його здійснення	7
1.2. Суб'єкти та об'єкти парламентського контролю	14
1.3. Законодавче окреслені дії суб'єктів парламентського контролю у залежності від об'єктів цього контролю	21
1.4. Питання ефективності парламентського контролю	28
<b>2. ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ПАРЛАМЕНТСЬКОГО КОНТРОЛЮ</b>	37
2.1. Правова інформація	39
2.2. Особливості правової інформації	44
2.3 Роль даних великих обсягів у праві	49
2.4. Особливості структури даних в праві	54
2.5. Складність автоматизованої обробки даних в праві	68
<b>3. КОНЦЕПЦІЯ BIG DATA В ПРАВІ</b>	70
3.1. Поняття Big Data	72
3.2. Техніки и технологи Big Data	77
3.2.1. Техніки Big Data	77
3.2.2. Технології Big Data	107
3.3. Джерела Big Data	118
3.4. Бази даних	129
3.5. Метадані в правовій інформатиці	137

<b>4. МЕРЕЖЕВІ МОДЕЛІ В ПРАВОВІЙ ІНФОРМАТИЦІ</b>	141
4.1. Концепція складних мереж	141
4.2. Мережі мови в правовій інформатиці	160
4.3. Корпусна лінгвістика в правовій інформатиці	170
4.4. Семантичні мережі, онтології	184
<b>5. ПОРІВНЯЛЬНИЙ АНАЛІЗ ДОКУМЕНТІВ</b>	197
5.1. Проблематика порівняльного аналізу документів	202
5.2. Алгоритми виявлення подібних документів	208
<b>6. ШТУЧНИЙ ІНТЕЛЕКТ У ПАРЛАМЕНТСЬКОМУ КОНТРОЛІ</b>	219
6.1. Поняття штучного інтелекту	219
6.2. Інтелектуальні агенти	228
6.3. Проблема контролю машинної обробки даних в праві	244
<b>ВИСНОВКИ</b>	254

## ПЕРЕДМОВА

Сукупний, у часі, вплив стрімкого науково-технічного прогресу, пандемії COVID-19, розв'язаної Російською Федерацією повномасштабної війни, а також цілеспрямованого забезпечення руху України у напрямку європейської та євроатлантичної інтеграції досить суттєво підвищили роль парламентського контролю, як одного з головних показників народовладдя в системі державного управління.

Парламентський контроль – складний і кропіткий процес, який вимагає участі значної кількості учасників, зокрема – експертів, вивчення та аналізу матеріалів нормативно правового, інформаційно-довідкового і науково-аналітичного характеру, статистичних та інших даних. Підсумкові рішення з парламентського контролю формуються і реалізуються парламентом на публічних засіданнях, але їх підготовка забезпечується роботою значної кількості суб'єктів парламентського контролю. Саме тому існують різні організаційні форми його діяльності, однією з основних яких є комітети і комісії.

На жаль, в силу різних факторів, як об'єктивних, так і суб'єктивних, оперативність, повнота та ефективність сучасного парламентського контролю досить суттєво не відповідають вимогам часу.

Як бачимо з наведеного, основна трудомісткість здійснення парламентського контролю полягає у зборі, систематизації та обробці матеріалів (інформації) для подальшого їх осмислення та первісної їх оцінки повноти та достовірності з метою формування попередніх висновків.

Саме тому, основною ціллю даного посібнику, є не тільки розгляд сучасного стану організації та проведення парламентського контролю, а розгляд шляхів забезпечення

відповідності парламентського контролю вимогам сучасності з урахуванням та використанням тих можливостей, які надають процеси цифровізації системи державного управління.

У даному посібнику були використані результати наукових досліджень науковців Державної наукової установи «Інститут інформації, безпеки і права Національної академії правових наук України», зокрема, за темами «Науково-методичне, правове та інформаційне забезпечення формування національної інтегрованої системи нормативно-правових актів в умовах децентралізації в Україні» під науковим керівництвом Коржа І.Ф., доктора юридичних наук, старшого наукового співробітника та «Організаційно-правові й технологічні засади розвитку системи нормативно-правової інформації та парламентського контролю в умовах цифровізації» під науковим керівництвом Ланде Д.В., доктора технічних наук, професора, а також наукового дослідження Кіндрата П.В. «Теоретичні засади оцінки ефективності законодавства в інформаційній сфері» на здобуття наукового ступеня «кандидат юридичних наук» під науковим керівництвом Фурашева В.М., кандидата технічних наук, доцента, старшого наукового співробітника.

# 1. ПАРЛАМЕНТСЬКИЙ КОНТРОЛЬ

## *1.1. Поняття «парламентський контроль» та нормативно-правова база його здійснення*

Здійснення будь-якого процесу, дії, в тому числі і процесів здійснення державного управління, у загальному вигляді, передбачає усвідомлення мети, нагальності та обсяжності, кінцевого результату та реальних або потенційних наслідків, прийняття остаточного рішення щодо здійснення, розробки алгоритму (послідовності дій) та забезпечення безумовного їх дотримання (контроль).

Під поняттями «контроль» та «парламентський контроль» будемо розуміти:

- контроль – це діяльність по спостереженню, аналізу та перевірці діяльності об'єктів контролю і спрямована на виявлення, запобігання та припинення дій, що суперечать чинному законодавству;

- парламентський контроль – одна із функцій діяльності Верховної Ради України.

Чому саме парламентський контроль відіграє суттєву роль в системі державного управління та забезпеченні життєдіяльності суспільства?

Стаття 5 Конституції України<sup>1</sup> чітко визначає, що:

- носієм суверенітету і єдиним джерелом влади в Україні є **народ**. Народ здійснює владу безпосередньо і через органи державної влади та органи місцевого самоврядування;

- право визначати і змінювати конституційний лад в Україні належить виключно народові і не може бути узурповане державою, її органами або посадовими особами;

---

<sup>1</sup> Конституція України: URL: <https://zakon.rada.gov.ua/laws/show/254к/96-вр#Text>

- **ніхто не може** узурпувати державну владу.

Також Конституція України у статті 6 встановила наступні складові системи державної влади: «Державна влада в Україні здійснюється на засадах її поділу на законодавчу, виконавчу та судову. Органи законодавчої, виконавчої та судової влади здійснюють свої повноваження у встановлених цією Конституцією межах і відповідно до законів України».

Відповідно до пункту 33) статті 85 здійснення парламентського контролю у межах, визначених Конституцією України та законом, віднесене до повноважень Верховної Ради України.

Вищенаведене досить чітко вказало на роль та місце парламентського контролю в системі державного управління.

Інша справа, як реалізуються ці положення Конституції України у реальності.

Як досить влучно відмітив український вчений Жилияев І.Б.<sup>2</sup>:

1) Основні функції парламенту: представницька, законодавча, контрольна закріплені конституційно. Однак, важко відокремити одну функцію від іншої – разом вони формують єдиний інститут, в якому кожна функція включає певні елементи іншої.

2) Парламентський контроль є одним із важливіших елементів державного контролю, він має бути тісно пов'язаним із системою державного контролю, застосовуючи свій оригінальний інструментарій (механізми та інструменти) контролю відповідно до загальних завдань розвитку українського парламентаризму.

---

<sup>2</sup> Жилияев І.Б. Інструментарій контрольних функцій парламентського комітету (на прикладі науки та вищої освіти). Інформація і право. № 4(43)/2022. – С. 139-148. URL: [http://ippi.org.ua/sites/default/files/15\\_24.pdf](http://ippi.org.ua/sites/default/files/15_24.pdf)



3) Правового визначення терміну «парламентський контроль» немає, що негативно впливає на детермінацію парламентського контролю. Використовується термін «контрольні функції парламенту», конкретного визначення змісту якого українським законодавством також не надано. Наукові дослідження використовують ці терміни у достатньо широкому змісті.

4) Основні елементи системи парламентського контролю відображені фрагментарно в окремих законах України. Комплексного уявлення про систему парламентського контролю, відповідної концепції розвитку української системи парламентського контролю не сформовано.

5) Функціональна система парламентського контролю побудована за принципами централізації – децентралізації, за якими кожний суб'єкт парламентського контролю має власні повноваження, що формують складну ієрархію суспільних відносин.

Виходячи із сутності положень Конституції України можна окреслити наступні **принципи** парламентського контролю:

- народовладдя;
- законності і доцільності контролю;
- визнання парламентського контролю вищою формою контролю в державі;
- врахування конституційного положення про те, що державна влада здійснюється на засадах її поділу на законодавчу, виконавчу та судову;
- гласності контролю;
- відповідальність суб'єкту контролю перед державою та підконтрольними об'єктами за об'єктивність контролю;

- систематичності, плановості, дієвості, неупередженості та ефективності контролю;

- виключного кола суб'єктів контролю;

- визнання в процесі контролю пріоритету особистих прав і свобод людини і громадянина, відокремлення громадянського суспільства від держави.

**Метою** парламентського контролю є формування цілісного і об'єктивного уявлення про стан правового забезпечення прав та свобод людини і громадянина, а також рівня правової урегульованості суспільних відносин, що виникають в сфері забезпечення життєдіяльності та сталого розвитку держави і суспільства з ціллю:

➤ встановлення відповідності діяльності підконтрольних об'єктів на території України або в межах її інтересів за кордоном Конституції і законам України; доцільності, соціально-економічної та політичної обґрунтованості їх діяльності;

➤ виявлення відхилень від конституційних засад та державних програм;

➤ визначення причин незаконної або недоцільної діяльності і засобів, спрямованих на удосконалення, раціоналізацію, розвиток діяльності об'єкту контролю;

➤ встановлення передумов щодо притягнення винних до відповідальності згідно з повноваженнями суб'єкта контролю.

Парламентський контроль - особливий вид суспільного контролю, який дає змогу парламентові отримувати достовірну й об'єктивну інформацію про фактичне виконання своїх обов'язків і повноважень відповідними органами влади та їх посадовими особами, а також оперативно реагувати на виявлені порушення.

Парламентський контроль, в першу чергу, спрямований на забезпечення конституційної законності та державної дисципліни у сферах їх безпосередньої владно-управлінської діяльності.

Необхідно підкреслити, що відповідно до ст. 24 Закону України «Про комітети Верховної Ради України»<sup>3</sup> рекомендації комітетів, які були висловлені за результатами розгляду об'єкта парламентського контролю, підлягають обов'язковому розгляду державними органами, органами місцевого самоврядування, їх посадовими особами, об'єднаннями громадян, підприємствами, установами і організаціями. Про результати розгляду та вжиті заходи має бути повідомлено комітетам у встановлений законом строк, якщо відповідними рекомендаціями не встановлено більш пізній строк.

Діяльність Верховної Ради України, в тому числі, здійснення парламентського контролю базується, в першу чергу, на положеннях Конституції України, зокрема, статей 75 – 101, та Віденській конвенції про право міжнародних договорів від 23 травня 1969 року, а також законів України:

- Про статус народного депутата України<sup>4</sup>;
- Про Регламент Верховної ради України<sup>5</sup>;
- Про комітети Верховної ради України<sup>6</sup>;

---

<sup>3</sup> Закон України Про комітети Верховної ради України: URL: <https://zakon.rada.gov.ua/laws/show/116/95-вр#Text>

<sup>4</sup> Закон України Про статус народного депутата України: URL: <https://zakon.rada.gov.ua/laws/show/2790-12#Text>

<sup>5</sup> Закон України Про Регламент Верховної ради України: URL: <https://zakon.rada.gov.ua/laws/show/1861-17#Text>

<sup>6</sup> Закон України Про комітети Верховної ради України: URL: <https://zakon.rada.gov.ua/laws/show/116/95-вр#Text>

- Про тимчасові слідчі комісії і тимчасові спеціальні комісії Верховної Ради України<sup>7</sup>;
- Про Уповноваженого Верховної Ради України з питань прав людини<sup>8</sup>;
- Про Рахункову палату<sup>9</sup>;
- Про звернення громадян<sup>10</sup>;
- Про інформацію<sup>11</sup>;
- Про доступ до публічної інформації<sup>12</sup>;
- Про державну службу<sup>13</sup>;
- Про засади державної антикорупційної політики на 2021-2025 роки<sup>14</sup>;
- Про запобігання корупції<sup>15</sup>;
- Кодексу законів про працю України<sup>16</sup> та спеціальними законами тощо.

---

<sup>7</sup> Закон України Про тимчасові слідчі комісії і тимчасові спеціальні комісії Верховної Ради України: URL: <https://zakon.rada.gov.ua/laws/show/400-20#Text>

<sup>8</sup> Закон України Про Уповноваженого Верховної Ради України з питань прав людини: URL: <https://zakon.rada.gov.ua/laws/show/776/97-вр#Text>

<sup>9</sup> Закон України Про Рахункову палату: URL: <https://zakon.rada.gov.ua/laws/show/576-19#Text>

<sup>10</sup> Закон України Про звернення громадян: URL: <https://zakon.rada.gov.ua/laws/show/393/96-вр#Text>

<sup>11</sup> Закон України Про інформацію: URL: <https://zakon.rada.gov.ua/laws/show/2938-17#Text>

<sup>12</sup> Закон України Про доступ до публічної інформації: URL: <https://zakon.rada.gov.ua/laws/show/2939-17#Text>

<sup>13</sup> Закон України Про державну службу: URL: <https://zakon.rada.gov.ua/laws/show/889-19#Text>

<sup>14</sup> Закон України Про засади державної антикорупційної політики на 2021-2025 роки: URL: <https://zakon.rada.gov.ua/laws/show/2322-20#Text>

<sup>15</sup> Закон України Про запобігання корупції: URL: <https://zakon.rada.gov.ua/laws/show/1700-18#Text>

Крім того, діяльність Верховної ради України регламентується постановами Верховної Ради України, зокрема, «Про перелік, кількісний склад і предмети видання комітетів Верховної Ради України поточного скликання», «Про структуру Апарату Верховної Ради України» та «Про План законопроектної роботи Верховної Ради на поточний рік» та відповідними розпорядженнями Голови Верховної Ради України.

Необхідно відмітити, що парламентський контроль Верховною Радою здійснюється в межах власної компетенції щодо будь-яких об'єктів контролю, за винятком судів та правоохоронних органів при провадженні ними оперативно-розшукової діяльності, дізнання та досудового слідства та здійснення правосуддя за окремими категоріями справ.

#### ***Питання для самоконтролю:***

- У зв'язку з чим система парламентського контролю має конституційне закріплення?
- Наведіть нормативно-правове визначення поняття «парламентський контроль».
- Наведіть основні принципи здійснення парламентського контролю.
- У чому полягає мета парламентського контролю?
- Основні цілі здійснення парламентського контролю.
- Які сфери діяльності які не підлягають парламентському контролю?

---

<sup>16</sup> Кодекс законів про працю України: URL: <https://zakon.rada.gov.ua/laws/show/322-08#Text>

## ***1.2. Суб`єкти та об`єкти парламентського контролю***

Частина 1 статті 1 Закону України «Про комітети Верховної Ради України» визначає: «Комітет Верховної Ради України – орган Верховної Ради України, який утворюється з числа народних депутатів України для здійснення за окремими напрямками законопроектної роботи, підготовки і попереднього розгляду питань, віднесених до повноважень Верховної Ради України, виконання контрольних функцій».

До завдань суб'єктів парламентського контролю в межах власної компетенції слід віднести:

1) здійснення контролю за виконанням Конституції і законів України, інших нормативних актів Верховної Ради України вищим, центральними та місцевими органами державної виконавчої влади і органами місцевого самоврядування, громадськими об'єднаннями, підприємствами, організаціями, установами незалежно від форм власності;

2) контроль законність і порядок надходження доходів в державний бюджет, законність, доцільність і ефективність витрат з державного бюджету;

3) вивчення документів відповідних об'єктів контролю, дослідження практики виконання функцій та повноважень органів, підприємств, організацій, установ, визначення тенденції розвитку в певних галузях державного та громадського життя;

4) формулювання пропозиції і рекомендації щодо удосконалення діяльності підконтрольного об'єкту, включаючи кадрові пропозиції.

Законом України «Про комітети Верховної Ради України» встановлюються три функції комітетів: 1) законопроектна, 2) організаційна, 3) контрольна.

Правова природа комітетів і комісій Верховної Ради України характеризується тим, що вони:

1) є складовою ієрархічної структури парламенту і обов'язковою ознакою парламентаризму;

2) є внутрішніми допоміжними органами Верховної Ради України;

3) визнаються суб'єктами контрольної діяльності парламенту;

4) їх членами можуть бути виключно народні депутати;

5) є колегіальними органами.

Як раніше вже було зазначено, суб'єктний склад парламентського контролю сформований ієрархічно та включає саму Верховну Раду України, її комітети і тимчасові спеціальні та слідчі комісії, народних депутатів, а також інших фізичних та юридичних осіб, спеціально уповноважених законом здійснювати окремі повноваження парламентського контролю

Законом України «Про комітети Верховної Ради України» головними суб'єктами здійснення парламентського контролю визначені комітети Верховної Ради України.

Суб'єкти парламентського контролю можуть уповноважити на виконання окремих контрольних завдань органи, організації, установи, окремих осіб, доручати їм також проведення наукових, юридичних, технічних і інших експертиз, надання висновків та пропозицій.

Відповідно до Закону України «Про комітети Верховної Ради України» основними **об'єктами** парламентського контролю визначено:

1) аналіз практики застосування законодавчих актів у діяльності державних органів, їх посадових осіб з питань, віднесених до предметів відання комітетів, підготовці та поданні відповідних висновків та рекомендацій на розгляд Верховної Ради України;

2) участь за дорученням Верховної Ради України у проведенні «години запитань до Уряду»;

3) контроль за виконанням Державного бюджету України в частині, що віднесена до предметів їх відання, для забезпечення доцільності, економності та ефективності використання державних коштів у порядку, встановленому законом;

4) організація та підготовка за дорученням Верховної Ради України парламентських слухань;

5) організації та підготовці слухань у комітетах;

6) підготовка та подання на розгляд Верховної Ради України запитів до Президента України від комітету відповідно до положень пункту 34 частини першої статті 85 Конституції України;

7) взаємодія з Рахунковою палатою;

8) взаємодія з Уповноваженим Верховної Ради України з прав людини;

9) направлення матеріалів для відповідного реагування в межах, установлених законом, органам Верховної Ради України, державним органам, їх посадовим особам;

10) розгляд на своїх засіданнях або під час слухань у комітеті звітів, доповідей та інформації державних органів та



посадових осіб, які, у передбачених законом випадках, подаються до Верховної Ради України, здійснення попередньої підготовки питань щодо розгляду на пленарному засіданні Верховної Ради України таких звітів, доповідей та інформації;

11) здійснення комітетом Верховної Ради України, до предмета відання якого віднесено питання забезпечення контрольних функцій Верховної Ради України за діяльністю органів спеціального призначення з правоохоронними функціями, правоохоронних органів спеціального призначення та розвідувальних органів, передбачених цим Законом заходів з метою гарантування неухильного і безумовного дотримання цими органами вимог Конституції України щодо забезпечення національної безпеки, недопущення їх використання для узурпації влади, порушення прав і свобод людини і громадянина.

Глава 3 «Права і обов'язки комітетів Верховної Ради України при здійсненні контрольної функції» Закону України «Про комітети Верховної Ради України» встановлює наступні **об'єкти** парламентського контролю:

- практика застосування державними органами, органами місцевого самоврядування, їх посадовими особами Конституції та законів України, аналіз відповідності закону прийнятих ними підзаконних нормативно-правових актів, своєчасності їх прийняття. За наслідками такого аналізу комітет вносить державним органам, органам місцевого самоврядування, їх посадовим особам рекомендації щодо приведення у відповідність із законом підзаконного нормативно-правового акта;

- рекомендації комітетів підлягають обов'язковому розгляду державними органами, органами місцевого самоврядування, їх посадовими особами, об'єднаннями

громадян, підприємствами, установами і організаціями. Про результати розгляду та вжиті заходи має бути повідомлено комітетам у встановлений законом строк, якщо відповідними рекомендаціями не встановлено більш пізній строк. (ст. 24);

- програми діяльності Кабінету Міністрів України, звіти про хід і результати їх виконання, звітів про виконання загальнодержавних програм економічного, науково-технічного, соціального, національно-культурного розвитку, охорони довкілля (п.1 ст. 25);

- питання відповідальності Кабінету Міністрів України та прийняття резолюції недовіри Кабінету Міністрів України ( п.2 ст. 25);

- затвердження рішень про надання Україною позик і економічної допомоги іноземним державам та міжнародним організаціям, а також про одержання Україною від іноземних держав, банків і міжнародних фінансових організацій позик, не передбачених Державним бюджетом України, здійснення контролю за їх використанням (п.3 ст. 25);

- оцінка стану дотримання та захисту прав і свобод людини в Україні на основі щорічної доповіді Уповноваженого Верховної Ради України з прав людини про стан дотримання та захисту прав і свобод людини в Україні та спеціальних доповідей Уповноваженого Верховної Ради України з прав людини (п.4 ст. 25);

- оцінка діяльності Генерального прокурора (п.5 ст. 25);

- оцінка виконання делегованих повноважень Верховною Радою Автономної Республіки Крим (п.6 ст. 25);

- оцінка діяльності на основі звітів, доповідей та інформації державних органів та посадових осіб, які у передбачених законом випадках подаються до Верховної Ради України (п.7 ст. 25).

Підсумовуючи викладені законодавчі положення щодо об'єктів парламентського контролю та спрямованості їх діяльності можемо констатувати наступне:

1. Об'єкти парламентського контролю — підконтрольні органи публічної влади (центральні органи виконавчої влади, органи місцевого самоврядування, їх посадові особи). Предмет парламентського контролю — дотримання об'єктом контролю відповідних правових норм: «Комітети з питань, віднесених до предметів їх відання, здійснюють аналіз практики застосування державними органами, органами місцевого самоврядування, їх посадовими особами Конституції та законів України, аналіз відповідності закону прийнятих ними підзаконних нормативно-правових актів, своєчасності їх прийняття».

2. До обов'язків об'єктів парламентського контролю необхідно віднести обов'язковість керівників органів державної влади та органів місцевого самоврядування, підприємств, організацій, установ повідомляти суб'єкт контролю про результати розгляду його запиту, надання суб'єктам парламентського контролю відомчих, локальних нормативних актів, сприяти роботі суб'єктів та їх представникам, особам, які залучаються до здійснення контрольних функцій.

3. В системі парламентського контролю можна виокремити два його види - прямого (безпосереднього) та непрямого (опосередкованого).

Об'єктами прямого парламентського контролю є:

- забезпечення конституційних прав, свобод та обов'язків людини і громадянина;

- діяльність Кабінету Міністрів України, генерального прокурора України, інших державних органів та посадових осіб, які обираються, призначаються чи затверджуються

Верховною Радою України або на призначення яких потрібна її згода;

- фінансово-економічної діяльності держави.

До об'єктів непрямого парламентського контролю слід віднести діяльність Президента України, зокрема під час:

- заслуховування щорічних та позачергових послань Президента про внутрішнє і зовнішнє становище України і затвердженні рішень про надання Україною позик і економічної допомоги іноземним державам та міжнародним організаціям, а також про одержання Україною від іноземних держав, банків та міжнародних фінансових організацій позик, не передбачених держбюджетом, здійсненні контролю за їх використанням;

- схвалення рішень глави держави про надання військової допомоги іншим державам, про направлення підрозділів Збройних сил України в іншу державу або про допуск підрозділів збройних сил інших держав на територію України;

- затвердженні Указів Президента України про введення воєнного або надзвичайного стану в Україні чи в окремих її місцевостях, про загальну або часткову мобілізацію, про оголошення окремих місцевостей України зонами надзвичайної екологічної ситуації;

- обговоренні подання Президента України про оголошення війни і укладення миру, рішення Президента України про використання Збройних сил України та інших військових формувань;

- наданні згоди на обов'язковість міжнародних договорів та денонсації міжнародних договорів України;

- прийнятті рішення про направлення до Президента України запиту на вимогу народного депутата, групи народних депутатів чи комітету Верховної Ради України;

- здійсненні процедури усунення Президента України з поста в порядку імпічменту.

***Питання для самоконтролю:***

- Суб'єкти парламентського контролю
- Основні завдання суб'єктів парламентського контролю
- Основні об'єкти парламентського контролю
- Предмет парламентського контролю
- Обов'язки об'єктів парламентського контролю
- Об'єкти прямого парламентського контролю
- Об'єкти непрямого парламентського контролю

***1.3. Законодавчо окреслені дії суб'єктів парламентського контролю у залежності від об'єктів цього контролю***

Суб'єкти парламентського контролю - комітети Верховної Ради України, у залежності від об'єкту цього контролю, відповідно до положень Закону України «Про комітети Верховної Ради України» повинні під час підготовки до розгляду Верховною Радою України здійснювати наступні дії:

*- під час підготовки до розгляду Верховною Радою України звіту Кабінету Міністрів України про хід і результати виконання схваленої Програми діяльності Кабінету Міністрів України* комітети розглядають на своїх засіданнях звіт Кабінету Міністрів України.

За результатами розгляду звіту Кабінету Міністрів України комітет приймає рекомендації. Такі рекомендації надсилаються комітету, визначеному головним з попереднього розгляду звіту Кабінету Міністрів України. Під час розгляду звіту Кабінету Міністрів України, підготовки рекомендацій Кабінету Міністрів України комітети

використовують інформацію, отриману ними під час реалізації контрольної функції, у тому числі при здійсненні аналізу практики застосування законодавчих актів, від Уповноваженого Верховної Ради України з прав людини, Рахункової палати, інших державних органів, органів місцевого самоврядування, їх посадових осіб, громадських об'єднань і громадян. (ст. 25-1);

- під час розгляду звітів, доповідей та іншої інформації державних органів та посадових осіб запрошувати на свої засідання осіб, участь яких є необхідною для розгляду цього питання; проводити відкриті для громадськості слухання у комітетах звітів, доповідей та іншої інформації державних органів та посадових осіб, поданих до Верховної Ради України (крім інформації з обмеженим доступом). Крім того, комітети зобов'язані розглядати на своїх засіданнях або під час слухань у комітеті звіти, доповіді та іншу інформацію державних органів та посадових осіб, які, у передбачених законом випадках, подаються до Верховної Ради України, а також здійснювати попередню підготовку питань щодо розгляду на пленарному засіданні Верховної Ради України звітів, доповідей та іншої інформації державних органів та посадових осіб, які у передбачених законом випадках подаються до Верховної Ради України, у порядку і строки, встановлені Регламентом Верховної Ради України. (ст. 25-2);

- під час здійснення контрольних функцій суб'єкти мають право на внесення пропозицій щодо бюджетної політики з урахуванням Бюджетної декларації та на виступ свого представника в обговоренні питання при розгляді проекту закону про Державний бюджет України в першому читанні, а також попередньо розглядають проект закону про Державний бюджет України на відповідний бюджетний період та не пізніше 1 жовтня року, що передує плановому, подають свої пропозиції щодо цього законопроекту до профільного

комітету та попередньо розглядають питання про виконання Державного бюджету України (включаючи висновки і пропозиції Рахункової палати щодо результатів контролю за дотриманням бюджетного законодавства) у частині, що віднесена до компетенції комітетів (ст. 27).

- під час підготовки і проведення парламентських слухань у Верховній Раді України суб'єкти парламентського контролю мають право за своїм рішенням ініціювати їх проведення у Верховній Раді України з метою вивчення питань, що становлять суспільний інтерес та потребують законодавчого врегулювання, стану виконання Кабінетом Міністрів України Конституції України, законів України, постанов Верховної Ради України та реалізують рішення комітету про їх проведення шляхом внесення народними депутатами України - членами комітету проекту постанови Верховної Ради України, підготовленого з дотриманням вимог Регламенту Верховної Ради України. Також готують висновки на аналітичні та довідкові матеріали, подані до Верховної Ради України Кабінетом Міністрів України, іншими державними органами, органами місцевого самоврядування та проекти постанов Верховної Ради України щодо схвалення відповідних рекомендацій за результатами парламентських слухань (ст. 28).

- під час проведення слухань у комітетах суб'єкти парламентського контролю мають право ініціювати пропозиції про проведення слухань, проводити їх за дорученням Верховної Ради України або згідно з планами роботи комітетів, а також проводити слухання з метою обговорення проектів найбільш важливих законодавчих актів, з'ясування ефективності реалізації прийнятих законів та інших актів Верховної Ради України з питань, віднесених до предметів їх відання, отримання всебічної інформації щодо питань, які розглядаються комітетом, їх детального вивчення

та обговорення, а також залучення широких кіл громадськості до участі у визначенні політики держави, розбудови демократичного суспільства. За результатами цих слухань суб'єкти парламентського контролю мають право використовувати інформацію, отриману під час слухань, при прийнятті рішень, наданні висновків, рекомендацій комітетом з питань, віднесених до предмета відання комітету, та розповсюджувати її серед народних депутатів України (ст. 29);

- *під час взаємодії суб'єктів парламентського контролю з Президентом України* Комітет може прийняти рішення про депутатський запит до Президента України і подати цей запит до Верховної Ради України для направлення в порядку, встановленому пунктом 34 частини першої статті 85 Конституції України та статтею 15 Закону України «Про статус народного депутата України» (ст. 30);

- *у взаємодії комітетів з Рахунковою палатою* суб'єкти парламентського контролю мають право розглядати узагальнені відомості, одержані відповідно до Закону України «Про Рахункову палату», а також розробляють на підставі доповідей та інформацій (повідомлень) Рахункової палати пропозиції щодо удосконалення законодавчих актів України, готують висновки з питань, передбачених Законом України «Про Рахункову палату» (ст. 31);

- *під час взаємодії комітетів з Уповноваженим Верховної Ради України з прав людини* суб'єкти парламентського контролю мають право на звернення до Уповноваженого Верховної Ради України з прав людини на підставі відомостей про порушення прав і свобод людини і громадянина з питань, віднесених до предмета відання комітету, органами державної влади, органами місцевого самоврядування та їх посадовими і службовими особами, а також на проведення аналізу щорічної та спеціальної



доповідей у Верховній Раді України Уповноваженого Верховної Ради України з прав людини та вносити пропозиції до проекту постанови Верховної Ради України за цими доповідями (ст. 32).

Відповідно до положень ст. 33-2 Закону України «Про комітети Верховної Ради України» здійснення суб'єктом парламентського контролю до предмета відання якого віднесено питання забезпечення контрольних функцій Верховної Ради України за діяльністю органів спеціального призначення з правоохоронними функціями, правоохоронних органів спеціального призначення та розвідувальних органів, крім визначених цим Законом прав та обов'язків комітетів Верховної Ради України базується на:

а) зібранні та аналізі інформації з питань, що належать до предмета його відання, а також питань, що стосуються національної та державної безпеки, а також інших надзвичайних подій, що можуть на них впливати;

б) отриманні від органів спеціального призначення з правоохоронними функціями, правоохоронних органів, правоохоронних органів спеціального призначення, органів, що реалізують державну політику у сфері цивільного захисту, захисту населення і територій від надзвичайних ситуацій та запобігання їх виникненню, та розвідувальних органів на постійній основі інформації стосовно загроз, що можуть впливати на національну та державну безпеку;

в) отриманні інформації від органів спеціального призначення з правоохоронними функціями, правоохоронних органів спеціального призначення та розвідувальних органів щодо надзвичайних подій, які сталися за участю співробітників таких органів;

г) проведенні закритих засідань з питань, віднесених до предмета відання комітету, та заслуховування на таких засіданнях керівників органів спеціального призначення з

правоохоронними функціями, правоохоронних органів спеціального призначення та розвідувальних органів та інших визначених комітетом осіб;

е) організації та здійснення підготовки за дорученням Верховної Ради України парламентських слухань з питань, віднесених до предмета його відання;

ж) участі у підготовці документів стратегічного планування у сферах діяльності органів спеціального призначення з правоохоронними функціями, правоохоронних органів спеціального призначення та розвідувальних органів;

з) заслуховуванні інформації про виконання органами спеціального призначення з правоохоронними функціями, правоохоронними органами спеціального призначення та розвідувальними органами покладених на них завдань;

і) розгляду звернення з питань, що віднесені до предмета його відання, у тому числі щодо рішень та дій органів спеціального призначення з правоохоронними функціями, правоохоронних органів спеціального призначення та розвідувальних органів;

ї) розгляду звіту за результатами здійснення Рахунковою палатою заходів державного зовнішнього фінансового контролю за використанням коштів Державного бюджету України органами спеціального призначення з правоохоронними функціями, правоохоронними органами спеціального призначення та розвідувальними органами, а також за дотриманням ними законодавства під час здійснення операцій з бюджетними коштами;

к) вивченням обставин і фактів, що наведені у матеріалах засобів масової інформації і несуть загрозу національній безпеці;

л) ініціюванням створення у встановленому законом порядку тимчасових слідчих комісій Верховної Ради України з питань, віднесених до предмета його відання;

м) проведенням попереднього обговорення кандидатур посадових осіб органів спеціального призначення з правоохоронними функціями, правоохоронних органів спеціального призначення та розвідувальних органів, яких згідно з Конституцією України обирає, призначає, затверджує Верховна Рада України або надає згоду на їхнє призначення;

н) оприлюдненням звітів про результати своєї діяльності, без зазначення в них інформації, що належить до таємної.

Необхідно підкреслити що, відповідно до положень ст. 57 Закону України «Про комітети Верховної Ради України» передбачена відповідальність за невиконання вимог цього Закону, а саме:

1. Посадові особи державних органів, органів місцевого самоврядування, підприємств, установ, організацій та об'єднань громадян за нез'явлення на засідання комітету без поважних причин, надання недостовірної, неповної інформації, відмову надати інформацію або за навмисне її приховування, створення штучних перешкод у роботі комітетів несуть відповідальність у встановленому законом порядку, крім випадків, передбачених законом.

2. За несвоєчасний розгляд рекомендацій комітетів та несвоєчасне повідомлення їм про вжиті заходи посадові особи державних органів, органів місцевого самоврядування, підприємств, установ, організацій та об'єднань громадян несуть відповідальність у встановленому законом порядку.

### ***Питання для самоконтролю***

Які законодавче окреслені дії суб'єктів парламентського контролю повинні відбуватися під час підготовки до розгляду Верховною Радою України звіту Кабінету Міністрів України про хід і результати виконання схваленої Програми діяльності Кабінету Міністрів України?

Які законодавче окреслені дії суб'єктів парламентського контролю повинні відбуватися під час розгляду звітів, доповідей та іншої інформації державних органів та посадових осіб?

Які права мають суб'єкти парламентського контролю під час здійснення контрольних функцій у сфері бюджетної політики?

Які права мають суб'єкти парламентського контролю під час підготовці і проведенні парламентських слухань у Верховній Раді України?

Які права мають суб'єкти парламентського контролю під час проведення слухань у комітетах?

Які права мають суб'єкти парламентського контролю під час взаємодії з Президентом України?

На чому базуються дії суб'єктом парламентського контролю до предмета відання якого віднесено питання забезпечення контрольних функцій Верховної Ради України за діяльністю органів спеціального призначення з правоохоронними функціями, правоохоронних органів спеціального призначення та розвідувальних органів?

#### ***1.4. Питання ефективності парламентського контролю***

Під час роботи Першої всеукраїнської науково-практичної конференції «Парламентський контроль в умовах децентралізації державної влади та цифрової трансформації в Україні: стан і проблеми» дійсний член Національної академії правових наук України В.Г. Пилипчук<sup>17</sup>, звертаючи увагу на проблемі парламентського контролю, акцентує увагу на

---

<sup>17</sup> Пилипчук В. Г. Системні проблеми та пріоритети розвитку парламентського контролю в контексті цифрових та суспільних трансформацій. Парламентський контроль в умовах децентралізації державної влади та цифрової трансформації в Україні: стан і проблеми : матеріали Першої всеукраїнської науково-практичної конференції, м. Київ, 30 березня 2021 р. / упор.: В. М. Фурашев, С. О. Дорогих. – Київ, 2021. – с. 12-20. URL: [http://ippi.org.ua/sites/default/files/zbirnik\\_tez\\_konferenciyi\\_\\_30.03.2021\\_0.pdf](http://ippi.org.ua/sites/default/files/zbirnik_tez_konferenciyi__30.03.2021_0.pdf)

проблему ефективності проведення багаточисленних парламентських чи комітетських слухань, розширених засідань профільних комітетів, заслуховувань представників міністерств та інших державних органів і прийнятих за результатами цих та інших заходів рішень. Він відносить до основних причин недоліків парламентського контролю: 1) обмеженість повноважень комітетів з питань парламентського контролю; 2) низьку ефективність наявних механізмів парламентського контролю (депутатських запитів, звітів уряду і представників державних органів тощо); 3) відсутність належної персональної відповідальності посадових осіб державних органів за невиконання чи неналежне виконання рішень Верховної Ради України та її профільних комітетів, прийнятих в рамках парламентського контролю. Він підкреслює, що проблема пошуку балансу між завданнями, функціями і повноваженнями різних гілок влади, належного врегулювання питань щодо завдань, повноважень та механізмів здійснення ПК залишається актуальною.

Під поняттям «ефективність парламентського контролю» будемо розуміти рівень впливу даного виду контролю на здатність держави забезпечувати за допомогою правових механізмів належний рівень збалансованості правотворчої, правозастосовної та правоохоронної діяльності з урахуванням інтересів людини, суспільства і держави.

Під ефективністю парламентського контролю розуміємо здатність суб'єкта парламентського контролю, або їх сукупності, впливати на суспільні відносини що виникають в правовій сфері системи державного управління внаслідок впровадження та забезпечення дотримання положень нормативно-правових актів і забезпечувати досягнення поставленої перед ними суспільно корисної мети.

Оцінювання ефективності парламентського контролю – суспільно-правове дослідження в рамках якого аналізується

спонукальна мотивація, цілі та завдання контролю, отримані результати, а також механізмів і наслідків їх впровадження.

**Метою оцінювання** ефективності парламентського контролю є формування цілісного і об'єктивного уявлення про ступінь досягнення цілей нормативно-правових актів та їх відповідності інтересам людини, суспільства і держави, а також встановлення причин пониження ефективності правового регулювання.

Основними **функціями оцінювання ефективності** є:

– оглядова (інформативна) – формує цілісне уявлення про функціонування чинної правової системи націленої на врегулювання досліджуваних відносин;

– аналітична – здійснює систематизацію і узагальнення реальних наслідків дії правових норм, розширюючи таким чином уявлення про особливості здійснення правового регулювання суспільних відносин, встановлення взаємозв'язків між законотворчою та законодавчою діяльністю і практикою застосування нормативно-правових актів;

– спрямовуюча – сприяє визначенню напрямків подальшого розвитку та вдосконалення чинного законодавства;

– коригуюча – визначає проблемні елементи предмету контролю що потребують редагування, видалення, заміни, або доповнення для забезпечення належного функціонування системи в цілому;

– моделююча – напрацювання достатнього масиву оцінок, що дозволяє розробляти моделі суспільної реакції на майбутні зміни предмету контролю.

Оцінка ефективності парламентського контролю - сукупності показників, які відображають повноту, своєчасність, об'єктивність та неупередженість застосування положень нормативно-правових актів, а також повноту, своєчасність, об'єктивність та неупередженість реагування на будь-які спроби не виконання/ігнорування цих положень.

Оцінка ефективності парламентського контролю – система показників, що відображають вплив різних аспектів його здійснення на рівень запобігання/усунення спонукальної мотивації проведення цього контролю.

Критерій ефективності – характеристика парламентського контролю з точки зору його впливу на державно-правові відносини і задоволення ними суспільних потреб.

Індикатор ефективності – параметри що відображають якісні характеристики парламентського контролю, зокрема обранням актуальності його спрямованості, організацією та проведенням, а також процесів і явищ викликаних впровадженням результатами його проведення.

Індикаторний показник – фактичні характеристики парламентського контролю, та зміни в соціальній, чи державній сфері, в наслідок його дії.

**Завданнями оцінки ефективності парламентського контролю є:**

- об'єктивне оцінювання об'єкту парламентського контролю у відповідності до предмету і сфери суспільних відносин, у відповідності до Конституції України, принципів побудови нормативно-правової системи, суспільних та державних інтересів;
- визначення елементів нормативно-правової системи та напрямків їх реалізації які знижують ефективність нормативно-правового врегулювання суспільних

відносин в сферах забезпечення життєдіяльності держави і суспільства, сталого їх розвитку;

- підготовка та прийняття обґрунтованого висновку щодо ефективності об'єкту, якій розглядається та, у разі необхідності, висунення пропозицій щодо шляхів та засобів її підвищення;
- встановлення відповідності мети і завдань об'єкта контролю суспільним потребам і напрямкам розвитку держави і суспільства;
- встановлення рівня ефективності через відповідно орієнтовну індикаторну систему.

### **Оцінювання ефективності парламентського контролю ґрунтується на принципах:**

1. законності – оцінювання ефективності має відбуватись в межах та з урахуванням вимог законодавчих актів;
2. наукової та практичної обґрунтованості – оцінювання мусить здійснюватися на основі теоретичних розробок та практичних напрацювань;
3. послідовності – оцінювання ефективності парламентського контролю мусить відбуватися у відповідності до продуманого і чітко визначеного плану здійснення контрольних функцій;
4. комплексності здійснення контрольних функцій – процес оцінювання потребує отримання й узгодження значної кількості показників та індикаторів з різних аспектів реалізації об'єкта парламентського контролю для формування якісної та об'єктивної оцінки;
5. об'єктивності та достовірності інформації – інформація що застосовуються для отримання оцінки ефективності має бути точною і підлягати верифікації;



6. експертності оцінок – для отримання й опрацювання інформації про суспільну значимість результатів дії законодавства необхідне залучення профільних спеціалістів, або формування експертної оцінки;

7. публічності результатів – результати та висновки оцінювання ефективності об'єкта парламентського контролю мають бути доступні широкому загалу;

8. відповідальності – суб'єкти оцінювання ефективності об'єкта парламентського контролю несуть відповідальність за надання неточної чи спотвореної інформації про оцінку її ефективності, або окремих її частин;

9. розумності (доцільності) пропозицій – рекомендації щодо напрямків підвищення ефективності об'єкта парламентського контролю мусять ґрунтуватися на глибокому розумінні суспільних явищ і процесів і уявлень про можливий ефект від зміни/трансформації об'єкта парламентського контролю. Якщо очікуваний ефект від змін не значний, то і пропонувати їх не варто;

10. плановості отримання оцінки – оцінювання ефективності об'єкта парламентського контролю доцільно здійснювати з визначеною періодичністю для аналізу змін в суспільному його сприйнятті та забезпечення підтримки ефективності.

### **Види оцінки ефективності**

Оцінка ефективності об'єкта парламентського контролю в залежності від стадії правотворчого процесу може поділятися на: перспективну і ретроспективну.

Перспективна оцінка ефективності здійснюється на різних етапах правотворчої роботи що передують введенню нормативно-правового акту в дію. В процесі перспективної оцінки визначальну роль відіграють показники що встановлюють якісні характеристики нормативно-правових

актів, а також відповідність поточним суспільним і державним інтересам.

Ретроспективна оцінка ефективності здійснюється після введення нормативно-правових актів в дію (зазвичай не раніше ніж через рік). Визначальними показниками даного типу оцінки виступають цілісність і дієвість правотворчості, рівень забезпеченості балансу суспільних і державних інтересів.

За підставами для проведення оцінка ефективності правотворчості може бути первинною, повторною, регулярною (систематичною).

Первинна оцінка ефективності проводиться вперше і покликана встановити вихідні положення для подальшого аналізу і порівняння складових правотворчості. Здійснюється для виявлення недоліків системи нормативно-правового забезпечення суспільних відносин, або для аналізу проекту нормативно-правового акту. Як наслідок вона може бути як перспективною, так і ретроспективною.

Повторна оцінка ефективності здійснюється після врахування висновків первинної оцінки і розробки проектів змін в об'єкт первинної оцінки. Повторна оцінка покликана встановити ефект від запропонованих змін, а тому має перспективний характер.

Регулярна (систематична) оцінка ефективності здійснюється згідно визначеного терміну і покликана встановити вплив змін у суспільстві та ефективність правотворчості, та виявити потребу у зміні чи вдосконаленні останнього. Має виключно ретроспективний характер.

За обсягом досліджень оцінка ефективності поділяється на цілісну і часткову.

Цілісна оцінка ефективності надає максимально розгорнутий висновок про ефективність оцінюваного законодавства, та враховує усю сукупність показників оцінки ефективності що можуть бути до нього застосовані.

Часткова оцінка ефективності покликана встановити ефективність окремих елементів, або напрямків впливу об'єкта парламентського контролю.

Підсумовуючи вищенаведене, можна відмітити що однією лише констатацією проблемних питань та труднощів, з якими доводиться стикатися під час організації та підготовці парламентських слухань, їх ефективність не підвищиться. Необхідно, використовуючи сучасні можливості засобів забезпечення оберт, збереження та обробки інформації із забезпеченням її безпеки, впровадження елементів штучного інтелекту та ін. розробляти та застосовувати нові прийоми і механізми максимального «усунення» людини під час здійснення найбільше трудомістких процесів/процедур організації та підготовці парламентських слухань.

Саме ці питання і розглядаються у **главі 2** даного посібника.

### ***Питання для самоконтролю***

- Що розуміємо під поняттям «ефективність парламентського контролю»?
- У чому полягає мета оцінювання ефективності парламентського контролю?
- Основні функції оцінювання ефективності.
- Складові оцінці ефективності парламентського контролю.
- Як розуміти поняття «критерій ефективності»?
- Що означає «індикатор ефективності»?
- Сутність поняття «індикаторний показник».

- Основні завдання оцінки ефективності парламентського контролю.
- Основні принципи на яких ґрунтується система оцінювання ефективності парламентського контролю.
- Види оцінки ефективності та коротка їх характеристика.

### **Рекомендована література до глави 1**

1. Консолідована система нормативно-правової інформації в умовах децентралізації державної влади в Україні – невід’ємна складова демократії : монографія / Корж І.Ф., Ланде Д.В., Лихоступ С. В. за заг. ред. Корж І. Ф. Київ : ТОВ “Видавничий дім “АртЕк”, 2021. 412 с.
2. Відкритість та прозорість в інформаційній діяльності Верховної Ради України: організаційно-правові аспекти : монографія / Дорогих С.О. Київ : Видавничий дім «АртЕк», 2018. 162 с.
3. Національна інтегрована система нормативно-правових актів: реальність і можливості /Корж І.Ф./Інформація і право.-2019.-№ 2.- с.194.- с.9-16; URL: <http://www.ippi.org.ua/korzh-if-natsionalna-integrovana-sistema-normativno-pravovikh-aktiv-realnist-i-mozhlivosti-st-9-16>
4. Парламентський контроль: практичні поради та рекомендації для підвищення ефективності : посібник / Інформаційне управління Апарату Верховної Ради України. URL: <http://iadm.rada.gov.ua/uploads/documents/68113.pdf>; <https://www.rada.gov.ua/news/Novyny/209096.html>

## 2. ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ПАРЛАМЕНТСЬКОГО КОНТРОЛЮ

Інформаційні технології парламентського контролю включають в себе використання різноманітних засобів та програмного забезпечення для збору, аналізу та інтерпретації даних, пов'язаних з діяльністю уряду, інших органів влади та організацій.

Застосування інформаційних технологій дозволяє ефективно контролювати дії уряду та забезпечувати виконання законів. Основні інструменти парламентського контролю, що застосовують інформаційні технології, включають в себе бази даних, електронні таблиці, системи обліку та моніторингу, інтернет-ресурси та електронні звіти.

Одним з важливих стандартів для інформаційних технологій парламентського контролю є стандарт Akoma Ntoso, що розроблений для стандартизації законодавчих документів та дозволяє їх електронне форматування та обмін між різними системами<sup>18</sup>. Створений Організацією розвитку стандартів структурованої інформації (OASIS) технічний комітет спрямовує свою роботу на Akoma Ntoso, що базується на мові XML та розвивається з 2004 р.<sup>19</sup> Спочатку розроблений для підтримки законодавчої діяльності парламентів африканських країн Akoma Ntoso наразі використовується іншими парламентами та інституціями й організаціями всього світу.

Для парламентського контролю можуть бути застосовані ІТ-стандарти, які спрямовані на забезпечення доступу до

---

<sup>18</sup> Електронна держава: нова ефективність урядування : монографія / А. А. Барікова. Київ : Юрінком Інтер, 2016. 224 с. ISBN 978-966-667-666-8.

<sup>19</sup> World e-Parliament Conference 2012 [Електронний ресурс]. Режим доступу : [http://www.ictparliament.org/sites/default/files/report\\_wepc\\_2012.pdf](http://www.ictparliament.org/sites/default/files/report_wepc_2012.pdf).

інформації та її захист від несанкціонованого доступу. Ось деякі з них:

ISO/IEC 27001 (Інформаційні технології. Методи захисту. Системи управління інформаційною безпекою. Вимоги) – цей стандарт встановлює вимоги до систем управління інформаційною безпекою (СУІБ) і надає фреймворк для оцінки та керування ризиками в галузі інформаційної безпеки.

ISO/IEC 27002 (Інформаційні технології. Технології безпеки. Практичні правила менеджменту інформаційної безпеки) – цей стандарт визначає практичні рекомендації з управління ризиками в галузі інформаційної безпеки та надає конкретні вказівки щодо реалізації заходів зі збереження інформації та захисту її від несанкціонованого доступу.

ISO/IEC 38500 (Корпоративне управління інформаційними технологіями) – цей стандарт встановлює принципи та вимоги до керування ІТ в організації, зокрема, управлінням ризиками та захистом інформації.

COBIT (Control Objectives for Information and Related Technology – Контрольні цілі для інформаційних та суміжних технологій) – це рекомендаційний фреймворк для управління ІТ-процесами, який надає детальний опис керуванням ризиками, контролю та управлінням змінами.

ITIL (Information Technology Infrastructure Library – бібліотека інфраструктури інформаційних технологій) – цей стандарт надає рекомендації щодо керуванням ІТ-інфраструктурою, процесами та послугами.

Застосування цих стандартів може допомогти забезпечити безпеку та доступ до інформації, збільшити ефективність та зменшити ризик порушення інформаційної безпеки при здійсненні парламентського контролю.

Застосування інформаційних технологій парламентського контролю та сучасних стандартів в галузі ІТ дозволяє підвищити ефективність роботи парламенту, сприяти більш швидкому та точному збору та обробці даних, забезпечувати більш транспарентну та відкриту діяльність уряду та інших органів влади.

## ***2.1. Правова інформація***

Стаття 17 Закону України «Про інформацію»<sup>20</sup> визначає, що «правова інформація - будь-які відомості про право, його систему, джерела, реалізацію, юридичні факти, правовідносини, правопорядок, правопорушення і боротьбу з ними та їх профілактику тощо».

Дана стаття також визначає що «джерелами правової інформації є Конституція України, інші законодавчі і підзаконні нормативно-правові акти, міжнародні договори та угоди, норми і принципи міжнародного права, а також ненормативні правові акти, повідомлення засобів масової інформації, публічні виступи, інші джерела інформації з правових питань».

Для здійснення парламентського контролю суто правової інформації недостатньо – потрібно мати різнобічну інформацію, яка повністю відображає інформаційну модель стану об'єкту парламентського контролю.

Як відоме, за порядком доступу інформація поділяється на відкриту та з обмеженим доступом . Обмеження права на одержання відкритої інформації не допускається.

---

<sup>20</sup> Про інформацію : Закон України від 13.01.2011 № 2938-VI. Відомості Верховної Ради України. 2011. № 32. С. 313.

Що стосується використання інформації з обмеженим доступом, то ст.21 Закону України «Про інформацію» встановлює наступне:

«1. Інформацією з обмеженим доступом є конфіденційна, таємна та службова інформація.

2. Конфіденційною є інформація про фізичну особу, а також інформація, доступ до якої обмежено фізичною або юридичною особою, крім суб'єктів владних повноважень. Конфіденційна інформація може поширюватися за бажанням (згодою) відповідної особи у визначеному нею порядку відповідно до передбачених нею умов, а також в інших випадках, визначених законом.

Відносини, пов'язані з правовим режимом конфіденційної інформації, регулюються законом.

3. Порядок віднесення інформації до таємної або службової, а також порядок доступу до неї регулюються законами.

4. До інформації з обмеженим доступом не можуть бути віднесені такі відомості:

1) про стан довкілля, якість харчових продуктів і предметів побуту;

2) про аварії, катастрофи, небезпечні природні явища та інші надзвичайні ситуації, що сталися або можуть статися і загрожують безпеці людей;

3) про стан здоров'я населення, його життєвий рівень, включаючи харчування, одяг, житло, медичне обслуговування та соціальне забезпечення, а також про соціально-демографічні показники, стан правопорядку, освіти і культури населення;

4) про факти порушення прав і свобод людини і громадянина;

5) про незаконні дії органів державної влади, органів місцевого самоврядування, їх посадових та службових осіб;

6) інші відомості, доступ до яких не може бути обмежено відповідно до законів та міжнародних договорів



України, згода на обов'язковість яких надана Верховною Радою України.»

Залежно від етапу парламентського контролю можна виділити первинну інформацію та вторинну, яка виникає в результаті оброблення первинної або іншої вторинної інформації. До вторинної належить інформація проміжна і результатна. Одержання результатної інформації є метою функціонування інформаційної системи (ІС).

З позиції технології ІС, у процесі розв'язування задач, розрізняють інформацію вхідну, проміжну і вихідну. Вхідною називають інформацію, що підлягає обробці – первинна і вторинна інформація, константи. Вихідна інформація є підсумком оброблення вхідних даних, але вона разом з результатною інформацією містить і деякі первинні дані. Проміжною є інформація, необхідна для розв'язування цих самих задач у наступних періодах.

За стабільністю інформація поділяється на постійну (сталу), яка не змінює своїх значень, умовно постійну, для якої це твердження може бути справедливим протягом тривалого періоду, та змінну, значення якої часто змінюються.

Різновиди інформації варто враховувати, організовуючи оброблення інформації, створюючи інформаційні системи, вибираючи варіанти технології розв'язування конкретних задач. Зазначимо також, правова інформація має бути вірогідною — об'єктивно відбивати реальність, повною — достатньою для розв'язування правової задачі, своєчасною — надходити до користувача тоді, коли той матиме в ній потребу.

Організація процесів отримання, використання, поширення та зберігання інформації, тобто інформаційної діяльності – сукупності дій, спрямованих на задоволення інформаційних

потреб громадян, юридичних осіб і держави, – істотно залежить від вибору носіїв та способу фіксації інформації на них. Одним з основних носіїв інформації, зокрема правової, є паперовий документ. Але з бурхливим зростанням обсягів правової інформації, яке почалося в розвинених країнах з другої половини ХХ століття разом з ускладненням завдань соціального управління і регулювання, постала невідкладна потреба в широкому використанні й інших носіїв. Сьогодні документ — це передбачена законом матеріальна форма одержання, зберігання, використання і поширення інформації фіксуванням її на папері, магнітній, кіно-, відео-, фотоплівці або на іншому носіїві. Це визначення стосується не тільки рукописних чи друкованих матеріалів на папері чи у вигляді книг, журналів, діаграм, карт тощо, а й матеріалів недрукованого походження (машино зчитуваних записів, фільмів, звукових записів) і тривимірних об'єктів чи реалій.

У контексті інформаційних систем і технологій особливого значення набуває таке поняття, як дані – інформація, подана у формалізованому вигляді, придатному для обробки автоматизованими засобами за можливої участі людини.

Сучасною формою організації даних на машинних носіях є автоматизовані банки даних. Автоматизований банк даних — це система інформаційних, математичних, програмних, мовних, організаційних і технічних засобів, необхідних для інтегрованого нагромадження, зберігання, ведення, актуалізації, пошуку та видачі даних. Основними складовими автоматизованого банку даних є база даних і система керування базою даних (СКБД).

База даних (БД) – це іменована структурована сукупність взаємозв'язаних даних, що відбиває стан об'єктів та відношень між ними в певній предметній галузі. БД призначається для використання багатьма користувачами у

процесі розв'язування кількох прикладних задач і не залежить від окремих прикладних програм. База даних перебуває під управлінням комп'ютерної системи управління базами даних (СКБД) — комплексу програмних і мовних засобів загального і спеціального призначення, необхідних для створення бази даних, підтримки її в актуальному стані, маніпулювання даними й організації доступу до них різних користувачів чи прикладних програм в умовах застосовуваної технології оброблення інформації.

Організація БД є необхідною передумовою для створення правових інформаційних систем і належного забезпечення правовою інформацією суспільства, але використання таких баз може призводити до нових проблем. Скажімо, нагромадження великого обсягу правової інформації в банку даних може призвести до монополізації, а згодом і до зловживань у вигляді приховування інформації, її незаконного оприлюднення чи використання з корисливою метою. Для запобігання таким зловживанням право власності на правову інформацію має належати державі, а використання даних регламентуватися законодавством.

***Питання для самоконтролю:***

- Який зміст має Закон України «Про інформацію»?
- Що входить до складу офіційної правової інформації та які є її джерела?
- Яка інформація з обмеженим доступом потрібна для здійснення окремих форм державної діяльності?
- Що відноситься до конфіденційної інформації?
- Які є джерела правової інформації?
- Які є основні носії інформації, зокрема правової?

- Як змінилися обсяги правової інформації у розвинених країнах з другої половини ХХ століття?
- Чому постала невідкладна потреба в широкому використанні й інших носіїв інформації?
- Що таке документ у контексті інформаційних систем і технологій?
- Що таке дані у контексті інформаційних систем і технологій?
- Що таке база даних і як вона відображає стан об'єктів та відношень між ними в певній предметній галузі?
- Що таке СКБД і яку роль вона відіграє у створенні та підтримці баз даних?
- Яка є необхідна передумова для створення правових інформаційних систем та забезпечення правовою інформацією суспільства?
- Які проблеми можуть виникнути при використанні баз даних, особливо при нагромадженні великого обсягу правової інформації?
- Як запобігти зловживанням при використанні баз даних, які містять правову інформацію?

## ***2.2. Особливості правової інформації***

Правова інформація — будь-які відомості про право, його систему, джерела, реалізацію, юридичні факти, правовідносини, правопорядок, правопорушення і боротьбу з ними та їх профілактику.<sup>21</sup> З іншого боку, правова інформація — це інформаційно-значеннєвий зміст правових норм: актів

---

<sup>21</sup> Інформація правова // Юридична енциклопедія : [у 6 т.] / ред. кол.: Ю. С. Шемшученко (відп. ред.) [та ін.]. — К. : Українська енциклопедія ім. М. П. Бажана, 1998. — Т. 2 : Д — Й. — 744 с. — ISBN 966-7492-00-8.

вищих та місцевих органів влади та управління, нормативні акти, судова, нотаріальна та арбітражна практика.<sup>22</sup>

Джерелами правової інформації є різноманітні законодавчі та підзаконні нормативні акти, зокрема Конституція України, а також міжнародні договори та угоди, які забезпечують норми та принципи міжнародного права. До інших джерел правової інформації можуть входити ненормативні правові акти, повідомлення засобів масової інформації та публічні виступи, а також інші джерела інформації з питань права.

Існує поділ джерел правової інформації на офіційні та неофіційні. До офіційних джерел правової інформації належать нормативні акти державних органів та акти вищих судових та арбітражних органів. Інші джерела є неофіційними, до яких можуть належати накази, вказівки керівників організацій, судові вироки, рішення та визначення, акти нотаріальних органів, науково-правові дослідження та юридичні доктрини.

Правова інформація може бути документованою або не документованою. Правовий документ – це будь-який матеріальний об'єкт, що містить різноманітні знання про право та призначений для передачі в часі і просторі та використання в суспільній практиці.

Традиційними матеріальними формами правового документа є паперовий документ, а сьогодні до них також відносять електронні носії інформації. Документи можуть бути традиційними (текстовими, графічними, аудіовізуальними) і мультимедійними (графічні, аудіо і відео файли).

---

<sup>22</sup> Міжнародна інформація. Кудрявцева С.П., Колос В.В., Навчальний посібник — К.: Видавничий Дім «Слово», 2005. - 400с.

Правова інформація має відповідати кільком критеріям. По-перше, вона повинна відображати реальні закономірності розвитку суспільства та бути пристосованою до історичних умов. По-друге, вона повинна точно відповідати чинному законодавству. По-третє, вона повинна бути повною, оскільки одне й те саме правове відношення може регулюватися різними нормативно-правовими актами. Повнота інформації є важливою передумовою для прийняття правильного рішення. По-четверте, вона повинна надходити до споживача вчасно. Як і будь-яка інша соціальна інформація, правова інформація є засобом відображення масових явищ, процесів та подій людської активності. Деякі властивості інформації є об'єктом правового регулювання<sup>23</sup>.

Правова інформація служить для відображення об'єктів, які можуть мати різні ознаки, що визначають їх здатність бути відображеними та уявлення про них. Ознаки можуть включати злочинні прояви, суб'єкти посягання, стадії вчинення, підстави та умови відповідальності, міри покарання та інше. У деяких законах визначається діяльність правоохоронних та інших державних органів щодо запобігання злочинним проявам. Більшість ознак об'єктів правової інформації не має чіткого зовнішнього вираження, тому їх потрібно визначати та з'ясовувати. Особливо це стосується інформації про якість (атрибутивна інформація), а не про кількість об'єктів чи їх ознак. Інформація також може деталізуватися, уточнюватися та змінюватися залежно від міри отримання відомостей про ці об'єкти.

Для формування уявлення про певний соціальний об'єкт необхідно враховувати інформацію про його прояви в різних

---

<sup>23</sup> Леонов Б.Д., Особливості правової інформації. Журнал «Інформація і право». –Т. 3(3), 2011. DOI: [https://doi.org/10.37750/2616-6798.2011.3\(3\).271640](https://doi.org/10.37750/2616-6798.2011.3(3).271640)

сферах, таких як економічна, управлінська, психологічна та інші. У випадку правової інформації, що стосується цих об'єктів, вона може мати економічне, управлінське, психологічне значення тощо, але в той же час вона також розкриває правову значущість цих об'єктів.

Однією з важливих властивостей інформації є її здатність до пізнання, інтерпретації та трансформації. Отримання правової інформації про певний об'єкт вимагає не тільки артибутивно-кількісної трансформації відповідних даних, але й спостереження, тлумачення та фахового аналізу<sup>8</sup>. Це особливо стосується інформації про терористичний акт, його мотиви, обставини, які призвели до його вчинення тощо. Якщо ми хочемо сформулювати загальне уявлення про стан певного соціального явища, процесу чи відносин, які складаються з багатьох подібних проявів, ми повинні задуматися про те, чи маємо ми достатню кількість таких проявів як безпосередніх об'єктів правової інформації, щоб отримати повне та вичерпне уявлення про цей об'єкт.

Існують різні методики узагальнення правової інформації з певних питань, такі як спостереження, правовий аналіз та нормативне групування. Хоча статистичні методи також можуть бути використані для узагальнення інформації, але самі статистичні показники не є об'єктом інформації. Важливо дотримуватися змісту та сутності інформації, і не допускати її викривлення.

Правова інформація має тісний зв'язок з управлінням, оскільки воно використовує багато термінів та понять, що описуються у законодавстві. Ця інформація має велике правове значення та є важливою для юристів.

На жаль, в законодавчому полі не існує окремого виду інформації, який би називався управлінською інформацією, яка використовується в сфері управління. Значна частина цієї

інформації входить до складу інших видів інформації, таких як дослідницька (аналітична), статистична та правова інформація. Часто ця інформація має правове значення, оскільки її джерелами є управлінські рішення уряду (доручення, розпорядження), міністерств, інших центральних органів виконавчої влади, правоохоронних органів або органів місцевого самоврядування. Інформаційно-аналітичні документи, матеріали перевірок та інша документація в галузі управління також містять управлінську інформацію.

Правова інформація загалом призначена бути засобом розв'язання багатьох правових питань. Головна мета правової інформації полягає в тому, щоб забезпечити повну та достовірну інформацію про закони, правила, процедури та процеси, пов'язані з правом. Іншими словами, мета полягає в забезпеченні правової впевненості та захисту прав і свобод кожної людини та суспільства в цілому. Досягнення цієї мети допомагає підтримувати правову державу та забезпечує взаємодію між всіма сторонами, що використовують правову інформацію.

***Питання для самоконтролю:***

- Які є джерела правової інформації?
- Що входить до офіційних та неофіційних джерел правової інформації?
- Що таке правовий документ?
- Які критерії повинна відповідати правова інформація?
- Які вимоги до її повноти та вчасності?
- Які сфери можуть мати значення для правової інформації щодо соціальних об'єктів?
- Які властивості інформації важливі для правового регулювання?
- Який зв'язок між правовою інформацією та управлінням?



- Чому управлінська інформація входить до складу інших видів інформації?

### ***2.3 Роль даних великих обсягів у праві***

Сучасні технології обробки даних великих обсягів (Big Data)<sup>24</sup> відіграють важливу роль у праві, оскільки вони забезпечують технологічну і інформаційну базу для прийняття рішень та виконання різних юридичних процедур.

Наприклад, у кримінальному праві, поліція та інші правоохоронні органи використовують дані великих обсягів для виявлення та розслідування злочинів. Дані можуть включати відеозаписи, електронні повідомлення, телефонні дзвінки та інші електронні докази, які можуть бути використані у судових процесах.

У цивільному праві, Big Data можуть використовуватися для розгляду справ, пов'язаних зі споживачами, належністю до власності, інтелектуальною власністю та іншими питаннями. Наприклад, компанії можуть використовувати дані про покупців для вирішення спорів щодо повернення товарів або збитків, пов'язаних зі зломом захисту даних.

Дані великих обсягів також можуть використовуватися для аналізу і прогнозування тенденцій у різних галузях права, таких як фінансове право, податкове право та регулювання бізнесу. Наприклад, у фінансовому праві, дані можуть використовуватися для виявлення шахрайства та маніпуляцій на ринку.

Дані великих обсягів є важливою складовою для аналізу тенденцій та прийняття рішень в галузі права з кількох причин.

---

<sup>24</sup> Оброблення надвеликих масивів даних (Big Data) : навчальний посібник. / Д.В. Ланде, І.Ю. Субач, А.Я. Гладун. / – Київ, ІСЗЗІ КПІ ім.. Ігоря Сікорського, 2021. – 168 с. ISBN 978-966-2344-83-7

По-перше, Big Data можуть містити значну кількість інформації, яка може допомогти виявити та проаналізувати закономірності та тенденції у різних сферах права. Наприклад, дані про кількість кримінальних справ за певний період можуть допомогти виявити тенденції щодо зростання або зменшення злочинності в певних регіонах, що може впливати на прийняття рішень про виділення ресурсів для поліції та інших правоохоронних органів.

По-друге, Big Data можуть допомогти в прогнозуванні та передбаченні майбутніх тенденцій у різних галузях права.

По-третє, Big Data можуть бути використані для вдосконалення систем правосуддя та прийняття рішень в конкретних випадках.

Дані великих обсягів можуть бути корисним інструментом для прогнозування та передбачення майбутніх тенденцій в парламентському контролі. Нижче наведено декілька прикладів того, як це можливо:

1. Аналіз даних про відвідування парламенту та участь у засіданнях може допомогти прогнозувати рівень зацікавленості громадян у роботі парламенту. Зокрема, якщо статистика показує зростання кількості відвідувань та участі у засіданнях, можна зробити висновок про зростання зацікавленості громадян у роботі парламенту.

2. Аналіз даних про кількість запитів на інформацію та звернень до парламентаріїв може допомогти виявити проблемні теми та питання, що вимагають уваги парламенту. Це може допомогти прогнозувати, які питання стануть актуальними у майбутньому та вимагатимуть уваги парламенту.

3. Аналіз даних про рівень довіри громадян до парламенту та парламентаріїв може допомогти прогнозувати рівень підтримки різних політичних партій та кандидатів на

майбутніх виборах. Наприклад, якщо довіра до однієї партії знижується, можна прогнозувати, що наступні вибори можуть стати складнішими для цієї партії.

4. Аналіз даних про кількість прийнятих законів та рішень може допомогти прогнозувати напрямки роботи парламенту в майбутньому.

5. Аналіз даних про голосування в парламенті може допомогти прогнозувати, які законопроекти та рішення можуть бути прийняті в майбутньому. Зокрема, якщо більшість парламентаріїв підтримує законопроект з певної галузі, можна прогнозувати, що цей закон буде прийнятий.

6. Аналіз даних про запити на інформацію та звернення до парламентаріїв може допомогти прогнозувати, які питання можуть стати актуальними в майбутньому та вимагатимуть уваги парламенту. Якщо збільшується кількість запитів на інформацію з певної теми, можна прогнозувати, що ця тема буде важливою у майбутньому.

7. Аналіз даних про рівень підтримки громадянського суспільства може допомогти прогнозувати, які законопроекти та рішення можуть бути підтримані громадянами. Зокрема, якщо опитування показує високий рівень підтримки громадянами певної ініціативи, можна прогнозувати, що ця ініціатива може мати високі шанси на підтримку в парламенті.

8. Аналіз даних про рівень корупції в політиці може допомогти прогнозувати, які питання можуть викликати скандали та потребувати особливої уваги з боку парламенту.

Дані великих обсягів можуть бути важливим інструментом для вдосконалення систем правосуддя та прийняття рішень в конкретних випадках, наприклад:

1. Аналіз даних про судову практику може допомогти зрозуміти, як судові рішення приймаються в певних випадках та як їх можна покращити. Аналіз даних про рішення відповідних судових органів у схожих випадках може допомогти зрозуміти, які фактори впливають на рішення суду та як їх можна врахувати у подальших судових процесах.

2. Використання аналітичних інструментів на основі даних великих обсягів може допомогти покращити ефективність системи правосуддя, зокрема, за допомогою автоматизованого аналізу доказів у судових процесах. Зокрема, за допомогою машинного навчання можна розробити алгоритми, які будуть прогнозувати висновки суду на основі аналізу великої кількості даних.

3. Аналіз даних про кримінальну статистику може допомогти вдосконалити систему превентивних заходів. Якщо дані показують зростання кількості злочинів певного типу у певних районах, можна прийняти превентивні заходи для зменшення кількості цих злочинів.

4. Використання аналітичних інструментів може допомогти приймати більш обґрунтовані та точні рішення в конкретних випадках, зокрема у справах з великою кількістю даних.

Інформаційне право, парламентський контроль та обробка великих даних пов'язані між собою з різних позицій.

З позиції інформаційного права, всі ці поняття пов'язані з обміном, зберіганням, доступом та захистом інформації.

У контексті парламентського контролю, інформаційне право розглядається як право громадян на отримання інформації від владних структур. Парламентський контроль передбачає, що інформація повинна бути доступна для депутатів та громадськості для забезпечення прозорості та демократії у діяльності влади.

Обробка Big Data зазвичай здійснюється з використанням персональних даних громадян. Тому при використанні великих даних необхідно дотримуватися норм інформаційного права, включаючи правила збору, зберігання та обробки персональних даних.

Інформаційне право, парламентський контроль та обробка великих даних є взаємопов'язаними поняттями, оскільки всі вони пов'язані з обміном, зберіганням, доступом та захистом інформації. Застосування цих понять в правовому контексті може допомогти забезпечити прозорість, демократію та захист прав громадян на інформацію.

***Питання для самоконтролю:***

- Яка роль сучасних технологій обробки даних великих обсягів у праві?
- Як поліція та інші правоохоронні органи використовують дані великих обсягів у кримінальному праві?
- Як компанії можуть використовувати дані великих обсягів у цивільному праві?
- Які галузі права можуть використовувати дані великих обсягів для аналізу і прогнозування тенденцій?
- Які можливості і переваги використання Big Data у галузі права?
- Дайте приклади їх застосування у парламентському контролі.
- Які можливості надає аналіз великих обсягів даних для вдосконалення систем правосуддя та прийняття рішень у конкретних випадках?
- Які зв'язки існують між інформаційним правом, парламентським контролем та обробкою великих даних?

- Які правила повинні дотримуватися при використанні великих даних, щоб забезпечити відповідність нормам інформаційного права?

## ***2.4. Особливості структури даних в праві***

Структура даних в праві відрізняється від структури даних в інших галузях знань, оскільки вона має свої власні особливості, які пов'язані з правовою природою інформації.

Особливості структури даних в праві полягають у наступному:

1. **Нормативність.** Дані в праві повинні бути нормативними, тобто мати правову силу та бути заснованими на законодавстві. Нормативність даних означає, що дані, що використовуються для парламентського контролю, повинні відповідати нормативним вимогам, тобто вони повинні бути отримані і оброблені відповідно до законодавства. Це важливо для забезпечення їх правової обґрунтованості та прийняття законних рішень на основі цих даних. Дані, що використовуються для парламентського контролю, можуть бути отримані різними способами, такими як запити на інформацію, звіти, аудити, аналізи тощо. Однак, незалежно від джерела, їх отримання та оброблення повинно відбуватися відповідно до вимог законодавства та принципів правової держави. Крім того, нормативність даних також передбачає, що дані повинні бути достовірними та мають бути підтверджені документально або іншими відповідними доказами. Такі дані забезпечують парламентаріїв можливістю приймати обґрунтовані та об'єктивні рішення на основі фактів та аналізів, що сприяє підвищенню ефективності та забезпеченню здійснення контролю відповідно до законодавства.

2. **Визначеність.** Дані в праві повинні бути точно визначені, оскільки невідомість або нечіткість можуть призвести до неправильних рішень та негативних наслідків. Визначеність даних означає, що дані, що використовуються для парламентського контролю, повинні бути чітко визначені та конкретні. Це дозволяє уникнути недостовірних та необ'єктивних даних, які можуть призвести до неправильних рішень. Визначеність даних також забезпечує зрозумілість та доступність інформації для парламентаріїв, що здійснюють контроль. Це дозволяє їм зрозуміти суть інформації та легше визначити, як вона може бути використана для прийняття рішень. Наприклад, якщо парламентарії здійснюють контроль над використанням бюджетних коштів, то дані, які використовуються для цього контролю, повинні бути чітко визначені та конкретні. Це можуть бути, наприклад, звіти про використання коштів, акти аудиту, інформація про укладені договори, деталі щодо виділених коштів на конкретні проекти тощо. Визначеність даних допомагає парламентаріям бути конкретними та цілеспрямованими у своїх запитах на інформацію та забезпечує достовірність і об'єктивність інформації, що використовується для контролю.

3. **Системність.** Дані в праві повинні бути системними, оскільки вони взаємодіють між собою та складаються з різних елементів, які мають свої відносини та залежності. Системність даних означає наявність логічного та організованого зв'язку між окремими даними, що дозволяє використовувати їх для формування комплексної інформації. Системність даних при парламентському контролі означає, що дані повинні бути організовані в систему, яка забезпечує можливість їх збору, зберігання, обробки та аналізу в масштабах, що необхідні для здійснення ефективного контролю. У контексті парламентського контролю системність даних відіграє важливу роль, оскільки дозволяє налагодити логічний та послідовний аналіз даних, отриманих

з різних джерел. Наявність системної структури даних дозволяє ефективніше здійснювати парламентський контроль, оскільки дозволяє зіставляти та аналізувати інформацію з різних джерел.

Наявність системної структури даних дозволяє ефективніше здійснювати парламентський контроль, оскільки дозволяє зіставляти та аналізувати інформацію з різних джерел.

Значення системності даних відображається у наступних аспектах:

- Ефективність контролю: Систематичний підхід до збору та обробки даних дозволяє парламентаріям здійснювати контроль ефективніше та швидше. Дані, які збираються, повинні бути організовані в логічні групи та підготовлені для аналізу. Це дозволяє швидше знайти потрібну інформацію та зрозуміти її суть.
- Об'єктивність та достовірність: Систематична обробка даних дозволяє забезпечити їх достовірність та об'єктивність. Система повинна забезпечувати точність та правильність даних, які збираються, а також захищати їх від можливої зміни чи впливу з боку третіх осіб.
- Системний підхід: Системний підхід до збору та аналізу даних дозволяє бачити всю картину здійснюваного контролю та зрозуміти, як різні частини процесу взаємодіють між собою. Це дозволяє бачити, які елементи контролю працюють ефективно, а які потребують покращення.

Таким чином, системність даних при парламентському контролі є важливим елементом, оскільки дозволяє



забезпечити комплексний аналіз та оцінку різних аспектів діяльності органів влади, що допомагає у прийнятті обґрунтованих та науково обґрунтованих рішень.

1. Нормативна логіка. Структура даних в праві має нормативну логіку, тобто вона повинна дотримуватися правил та законів, які визначають взаємозв'язки між різними елементами. Нормативна логіка має велике значення при парламентському контролі, оскільки це допомагає забезпечити дотримання норм та правил у законодавчій діяльності. Нормативна логіка базується на принципах логіки та наук про право, що забезпечує консистентність та узгодженість законодавчої діяльності зі стандартами та нормами. В контексті парламентського контролю, нормативна логіка допомагає розуміти та інтерпретувати законодавчі акти, розробляти стратегії взаємодії з іншими учасниками законодавчого процесу, визначати правові наслідки різних варіантів дій та приймати рішення на основі чіткого розуміння норм та їх застосування.

Консистентність та узгодженість законодавчої діяльності зі стандартами та нормами важлива для забезпечення правової стабільності та довіри до правової системи. Це означає, що законодавство повинне бути написано таким чином, щоб не суперечити існуючим нормам та стандартам, а також іншим законам та рішенням.

Наприклад, якщо закон про захист персональних даних суперечить загальним стандартам захисту персональних даних, то це може призвести до проблем при застосуванні закону. Крім того, якщо законодавча діяльність не відповідає загальним стандартам та нормам, то це може підірвати довіру громадян до правової системи та держави загалом.

Тому важливо, щоб законодавча діяльність була консистентною та узгодженою з існуючими стандартами та

нормами. Це можна досягти шляхом системного аналізу законодавства, врахування міжнародних зобов'язань та принципів, забезпечення єдності та цілісності законодавства.

Консистентність – це властивість системи або процесу, яка характеризується внутрішньою згодою, взаємодією та відповідністю між різними її складовими. У контексті права консистентність вимагає, щоб закони та інші нормативні акти не суперечили одне одному та не порушували конституційні права та свободи людини. Консистентність також передбачає відповідність рішень та дій державних органів загальному правовому порядку і законодавчим стандартам. Важливість консистентності в праві полягає в тому, що вона допомагає забезпечити стабільність правової системи та виключити можливість виникнення суперечностей та протиріч у законодавстві, що можуть призвести до несправедливості, правової нестабільності та невизначеності. Також консистентність сприяє захисту прав та інтересів людей, оскільки вона забезпечує прозорість та передбачуваність дій держави та її органів.

Для забезпечення нормативної логіки при парламентському контролі, важливо мати доступ до актуальної та точної інформації про законодавчі акти, їх зміни та взаємозв'язки з іншими нормами та правилами. Також важливим є розуміння основних принципів логіки та наук про право, що дозволяє правильно аналізувати та інтерпретувати дані та приймати обґрунтовані рішення.

2. Ієрархічність. Структура даних в праві може мати ієрархічну форму, оскільки законодавчі акти можуть бути розміщені на різних рівнях: від Конституції до звичайного закону. Ієрархічність даних в парламентському контролі означає, що дані систематизовані у вигляді ієрархії, тобто розташовані в логічному порядку від загального до деталей. Це допомагає вести звітність із застосування нормативних

актів та ухвалених рішень в парламенті. Ієрархічна структура даних дозволяє легко організувати та зберігати дані, що стосуються законодавства та правової практики, забезпечує зручний та швидкий доступ до цих даних та сприяє аналізу законодавчої діяльності. Також ієрархічність даних дозволяє враховувати ієрархічну структуру законодавства та відповідних документів, що сприяє їх зрозумінню та ефективному застосуванню. Крім того, ієрархічна структура даних забезпечує зручність при веденні звітності та контролі за дотриманням законодавства та відповідних процедур, що робить її важливим елементом парламентського контролю.

3. Формальність. Структура даних в праві повинна бути формальною, оскільки вона має відповідати певним формальним вимогам, які забезпечують її правову силу та вплив на поведінку людей. Відповідність даних певним формальним вимогам є дуже важливим аспектом у парламентському контролі. Це означає, що дані повинні бути зібрані, організовані та представлені у форматі, який відповідає стандартам та вимогам, встановленим законодавством.

Наприклад, у багатьох країнах законодавство вимагає, щоб певна інформація була надана у певному форматі, такому як структурований формат даних XML. Якщо дані не відповідають цим вимогам, вони можуть бути відхилені, що може призвести до невдалих результатів парламентського контролю.

Однак, важливо зазначити, що відповідність формальним вимогам не завжди гарантує якість даних та їхню правильність. Тому, крім відповідності формальним вимогам, також необхідно забезпечити якість та достовірність даних.

4. Абстрактність. Структура даних в праві може бути абстрактною, оскільки вона не завжди пов'язана з

конкретними фактами чи обставинами, а може бути визначена загальними поняттями та категоріями. Абстрактність структури даних в парламентському контролі дозволяє створювати загальні моделі та правила для опису різних аспектів законодавства, що забезпечує їх універсальність та застосовність для різних випадків та обставин. Це допомагає забезпечити однаковий підхід до розгляду різних законопроектів та забезпечити консистентність законодавства в цілому. Наприклад, створення загальних форматів для подання даних, що містяться в законопроектах, дозволяє автоматично аналізувати ці дані та порівнювати їх з іншими даними, що забезпечує більш ефективний та об'єктивний аналіз та прийняття рішень. Крім того, абстрактність структури даних дозволяє підвищити рівень автоматизації процесів парламентського контролю, наприклад, створення автоматизованих систем для аналізу великих обсягів даних з метою виявлення тенденцій та прогнозування результатів рішень.

Розвиток інформаційних технологій передбачає постійне зростання впливу використання цифрових технологій в різноманітних сферах суспільного життя, в тому числі в усьому, що так чи інакше пов'язане з генерацією, обробкою та зберіганням інформації. Складність зокрема полягає в тому, що комп'ютерні системи вимагають, щоб інформація, з якою вони працюють, була належним чином формалізована. А цю вимогу важко реалізувати в умовах несинхронізованої діяльності багатьох незалежних агентів інформаційних процесів (термін «агент» в нашому випадку використовується в широкому розумінні, і ми не будемо наводити його строге визначення). Маємо на увазі організації та фізичні особи, які приймають участь в формуванні правової інформації.

Закон України «Про інформацію»<sup>25</sup> визначає її як «будь-які відомості та/або дані, які можуть бути збережені на матеріальних носіях або відображені в електронному вигляді». Для нас важливим є те, що інформація визначається як відомості або дані. Зупинімось на цьому моменті детальніше. Поняття «відомості» має різні значення, але в будь-якому разі вони є чимось таким, що виникає в людській свідомості і тому не може сприйматися то оброблятися машиною, в тому числі зберігатись на матеріальних носіях та відображатись на електронних пристроях. Навпаки, поняття «дані» визначається як «формалізоване подання інформації, придатне для інтерпретування, пересилання чи оброблення за участю людини або автоматичними засобами», тобто мають суто об'єктивну природу. Але саме відомості сприймає кінцевий споживач. Тому актуальною є проблема відношення між відомостями та даними. Зберігаються та обробляються машиною дані, а людиною сприймаються відомості.

І це є надзвичайно важливим в галузі правової інформації, оскільки в правовій сфері потрібні не загальні враження та уявлення, а чіткі формулювання. Нагадаємо, що права інформація - це «будь-які відомості про право, його систему, джерела, реалізацію, юридичні факти, правовідносини, правопорядок, правопорушення і боротьбу з ними та їх профілактику тощо».

Із наведеного визначення видно, що права інформація як така має досить складну структуру. Ця складність посилюється неоднорідністю її джерел, до яких відносяться: «Конституція України, інші законодавчі і підзаконні нормативно-правові акти, міжнародні договори та угоди,

---

<sup>25</sup> Про інформацію : Закон України від 02.10.1992 № 2657-XII. Відомості Верховної Ради України. 1992. № 48. С. 650.

норми і принципи міжнародного права, а також ненормативні правові акти, повідомлення засобів масової інформації, публічні виступи, інші джерела інформації з правових питань». Отже, машинна обробка таких складних за своєю структурою масивів даних, яка забезпечує правильне інтерпретування, є непростотою та відповідальною справою, що породжує низку проблем як теоретичного, так і технічного характеру.

Дані в праві можуть бути розподілені на ряд масивів залежно від рівня структурованості та формальних характеристик структури, включно з повністю неструктурованими текстами, які важливі для розуміння інших, структурованих даних.

Прикладом добре структурованої інформації можуть служити реляційні бази даних. Дані в них організовані таким чином, що до кожного інформаційного блоку, що має визначений смисл, надається прямий доступ. Він завжди може бути реалізований шляхом інформаційного пошуку за запитом, побудованим в рамках технології SQL. Такі запити можуть вводитися користувачами для задоволення власних інформаційних потреб, або генеруватися програмними комплексами для виконання внутрішніх робіт з даними. А це дозволяє, в свою чергу, будувати системи довільної складності формальних зв'язків (реляцій) між цими блоками. Тому реляційні бази даних, а також довільні їх комплекси завжди можуть бути однозначно представлені певною мережею (як правило, одноранговою). З такими даними, якщо вони вже існують, дуже легко працювати, але створення і супровід реляційних баз даних є складним процесом, що вимагає значних ресурсів, в тому числі фінансових.

Як приклад неструктурованої інформації наведемо стенограми різноманітних виступів правників, що містять

тлумачення та інтерпретації правових норм. Вони, взагалі кажучи, являють собою документальні колекції, в яких кожний документ має власний формат (або взагалі не має його) і в якому відсутнє формальне виділення окремих смислових блоків. Через відсутність формальних зв'язків між документами такі набори даних або не можуть бути представлені мережею, або мережа є тривіальною.

Робота з структурованими і неструктурованими даними якісно відрізняються між собою. Інформаційний пошук в реляційних базах даних передбачає побудову наборів значень полів, що відповідають умовам запиту, тоді як повнотекстовий пошук, що використовується для роботи з документальними масивами полягає в побудові вибірки документів, які містять в собі задані слова та словосполучення. Також у випадку неструктурованих даних актуальними є специфічні задачі, такі як автоматичне реферування документів, створення дайджестів, побудова систем інформаційних зв'язків, вибіркоче поширення інформації, тощо.

Останніми роками надзвичайно поширеною стала технологія створення і обробки документальних колекцій, документи в яких побудовані з використанням тієї чи іншої мови розмітки (наприклад, XML). Ця технологія є до певної міри проміжною, але в багатьох випадках значно спрощує вирішення практичних задач, оскільки, з одного боку, дозволяє створювати досить розвинені структури, в тому числі семантичні. А з другого боку, вимагають відносно прості і не затратні інструментальні засоби обробки даних.

В будь-якому випадку при машинній обробці програмний комплекс повинен працювати одночасно з даними різної структури.

Очевидно, що в умовах цифрової трансформації зростає і надалі зростатиме рівень автоматизації відповідних процесів. При цьому зазначимо, що про повний перехід до автоматичних систем обробки змістовної інформації в найближчому історичному майбутньому не йдеться, оскільки для цього необхідно достатньою мірою вирішити проблему створення штучного інтелекту. Тому на даному етапі технологічного розвитку актуальним є використання автоматизованих систем. Нагадаємо, що автоматизована система (на відміну від автоматичної) не заміняє людину, а спрощує її працю шляхом передачі певної кількості функціональних можливостей машині. Тому автоматизація виробничих процесів за своєю метою передбачає зменшення обсягів людської (ручної) праці. Але людську працю можна умовно розділити на дві категорії. Перша – це власне виконання виробничих операцій. А друга – підготовчі операції, необхідні для функціонування апаратно-програмного комплексу. До цієї категорії відноситься підготовка вхідних даних відповідно до вимог комп'ютерної системи (агрегування, редагування, форматування тощо). Такі операції можуть бути достатньо трудомісткими і потребувати багато часу та ресурсів, включно з робочою силою. Тому здійснення ефективної автоматизації вимагає максимального скорочення підготовчих робіт.

Тому особливої актуальності в першу чергу набуває проблема забезпечення належної уніфікації структури правової інформації, яка може бути визначена як «процес приведення чинного права до єдиної системи, усунення розбіжностей і надання одноманітності правовому регулюванню подібних або близьких видів суспільних відносин»<sup>26</sup>. Уніфікація потрібна в першу чергу для того, щоб

---

<sup>26</sup> Сучасне загальноєвропейське кримінальне законодавство: проблеми гармонізації : монографія / М. І. Хавронюк. – Київ: Істина, 2005. - 264 с.



надати, наскільки це можливо, технології обробки та зберігання правової інформації універсального характеру. Це дозволить максимально автоматизувати процедури підготовки текстів до завантаження в апаратно-програмні комплекси всіх організацій, які задіяні в загальному процесі. Універсальність також гарантує правильність інтерпретування даних, що підлягають обробці і використанню.

Практично, уніфікація містить такі основні компоненти:

- мінімізація передбачених вихідних форм документів;
- відповідність документів визначеним правилам оформлення;
- зберігання документів у відповідності з узгодженим переліком визначених типів даних, структур інформаційних одиниць та системи зв'язків між ними;
- використання єдиного комплексу класифікаторів, рубрикаторів, тезаурусів, довідників;
- використання єдиного набору реквізитів документів для кожного передбаченого типу;
- стандартизація наборів метаданих документів для кожного передбаченого типу;
- введення даних з використанням єдиного комплексу спеціалізованих технічних засобів;
- здійснення модифікації та видалення документів в рамках єдиного комплексу технічних умов;
- виведення вихідної інформації у формі, прийнятній для непідготовлених користувачів;
- можливість обміну даними через локальні, корпоративні та глобальні мережі.

Реалізація цих компонентів передбачає принаймні три основні напрями:

- напрацювання нормативної бази (посадові інструкції тощо);

- створення інформаційної бази (розробка класифікаторів тощо);
- розробка технічних засобів (конвертори тощо).

Уніфікація структури правової інформації має стати невід'ємною складовою комплексної програми розвитку системи нормативно-правової інформації та парламентського контролю.

Але, нажаль, повна уніфікація далеко не завжди може бути досягнута. Головну роль відіграють дві причини. Перша – технічна. Для повної уніфікації навіть одного, окремо взятого інформаційного комплексу необхідні значні ресурси, що включають в себе, окрім іншого, витрати часу та використання колективів кваліфікованих працівників. Друга причина – організаційна. Потрібно забезпечити виконання таких робіт, а це досить складна задача, в тому числі через те, що уніфікація передбачає не лише разове створення певної інформаційної системи, але й постійного її супроводу.

Тому особливості структури інформації в праві є важливим чинником, який завжди слід враховувати при розробці та здійсненні різноманітних інформаційно-технологічних проєктів.

***Питання для самоконтролю:***

- Яка особливість структури даних в праві порівняно з іншими галузями знань?
- Що означає нормативність даних в праві?
- Яким чином визначеність даних впливає на контроль діяльності в праві?
- Які способи отримання та обробки даних використовуються для парламентського контролю?
- Як відповідність даних вимогам законодавства забезпечує об'єктивність прийнятих рішень?

- Що означає системність даних?
- Яку роль відіграє системність даних у парламентському контролі?
- Як систематична обробка даних дозволяє забезпечити їх достовірність та об'єктивність?
- Як системний підхід до збору та аналізу даних допомагає розуміти, як різні частини процесу контролю взаємодіють між собою?
- Як нормативна логіка допомагає в контексті парламентського контролю?
- Що таке консистентність в контексті права і чому вона важлива для правової системи?
- Чому консистентність та узгодженість законодавчої діяльності зі стандартами та нормами є важливою для правової стабільності та довіри до правової системи?
- Які можуть бути наслідки, якщо законодавство суперечить існуючим нормам та стандартам?
- Як можна досягти консистентності та узгодженості законодавчої діяльності з існуючими стандартами та нормами?
- Які переваги має ієрархічна структура даних в праві та парламентському контролі?
- Як ієрархічна структура даних сприяє збереженню та організації законодавчих даних?
- Як відповідність даних формальним вимогам впливає на їх правову силу та ефективність парламентського контролю?
- Які складові має включати комплексна програма розвитку системи нормативно-правової інформації та парламентського контролю?
- Чому повна уніфікація структури правової інформації є складною задачею?

- Які ресурси необхідні для повної уніфікації окремо взятого інформаційного комплексу?
- Які особливості структури інформації в праві важливі при розробці та здійсненні інформаційно-технологічних проектів?
- Як можна забезпечити постійний супровід інформаційної системи після її створення?

### ***2.5. Складність автоматизованої обробки даних в праві***

Автоматизована обробка даних в праві може бути дуже складною з різних причин. Нижче наведено кілька причин:

1. **Різноманітність даних:** в правовій галузі може бути багато різних типів даних, таких як текстові документи, електронні таблиці, бази даних, відео- та аудіо записи тощо. Всі ці дані можуть бути потрібні для вирішення конкретних правових питань, і обробка їх усіх може бути дуже складною.
2. **Низька стандартизація:** правова галузь може мати різні стандарти для організації та представлення даних, що може ускладнити автоматизовану обробку даних. Наприклад, формати документів, способи кодування даних тощо можуть різнитися в залежності від країни або регіону.
3. **Конфіденційність даних:** у правовій галузі можуть бути конфіденційні дані, такі як особиста інформація про клієнтів, розгляди судових процесів тощо. Забезпечення конфіденційності даних може ускладнити автоматизовану обробку даних.
4. **Потреба у точності:** у правовій галузі потрібна висока точність в обробці даних. Навіть невеликі помилки можуть мати серйозні наслідки для

рішення судових справ або прийняття рішень урядових структур.

5. Великий обсяг даних: у правовій галузі може бути великий обсяг даних, особливо великі організації або урядові структури. Обробка таких великих обсягів даних може бути важкою та ресурсоємною задачею.

Тобто дані в праві мають специфічну структуру та вимагають певного рівня точності та нормативності, що ускладнює їх автоматичну обробку. Крім того, існує проблема недостатньої якості та стандартизації даних в правовій сфері. Для боротьби зі складністю автоматизованої обробки даних в праві необхідно використовувати спеціалізовані програмні засоби та технології, які дозволяють проводити аналіз та обробку даних відповідно до нормативних вимог. Також важливо створювати єдині стандарти та формати для зберігання та передачі даних в правовій сфері, що сприятиме їх більш ефективній автоматизованій обробці.

***Питання для самоконтролю:***

- Які типи даних можуть бути необхідні для вирішення конкретних правових питань у правовій галузі?
- Які проблеми можуть виникнути через різні стандарти для організації та представлення даних у правовій галузі?
- Які конфіденційні дані можуть зустрічатися у правовій галузі та як їх можна захистити під час автоматизованої обробки даних?
- Яка роль точності у правовій галузі та які можуть бути наслідки невірної обробки даних?
  - Які проблеми можуть виникнути через великий обсяг даних у правовій галузі та як їх можна вирішити?

### 3. КОНЦЕПЦІЯ BIG DATA В ПРАВІ

На цей час у правовій інформатиці дедалі більшу роль грає поняття «великих даних» (Big Data)<sup>27</sup>. Фахівці, які займаються обробкою, агрегацією, аналізом масивів даних, вирішенням проблем, зумовлених зростанням їх обсягів, динамікою, варіативністю, нині називають «вченими з обробки даних» (Data Scientists), відповідно, наука – Data Science<sup>28</sup>.

Чому обсяги даних перетворилися на проблему? Проблема полягає в тому, що збільшення обсягів даних, що збираються, зберігаються і обробляються, вимагає відповідних інструментів та технологій для ефективної обробки та аналізу цих даних. Застосування традиційних інструментів управління базами даних та додатків для обробки таких об'ємів даних може бути недостатнім та не забезпечувати необхідну швидкість та точність обробки. Більш того, із збільшенням обсягів даних зростає складність їх зберігання, обробки та аналізу, тому що дані можуть бути неструктурованими та містити складні зв'язки та взаємодії. У разі неправильного або недостатнього аналізу таких даних можуть бути недостовірні та використані для неправильних рішень. Отже, проблема полягає в тому, що потрібні нові інструменти та технології для ефективного зберігання, обробки та аналізу даних в масштабах Big Data.

Незважаючи на те, що комп'ютери ставали все швидкодіючими, а розміри пам'яті та обсяги даних зростали

---

<sup>27</sup> Big Data Analytics: A Social Network Approach. Mrutyunjaya Panda, Aboul-Ella Hassanien, Ajith Abraham. – CRC Press, 2018. – 322 pp.

<sup>28</sup> Introducing Data Science: Big Data, Machine Learning, and more, using Python tools. Davy Cielen, Arno Meysman, Mohamed Ali. – Manning Publications, 2016. – 320 pp.

ще швидше, лише деякі алгоритми можуть лінійно масштабуватися зі зростанням вхідних даних. Основні проблеми пов'язані зі збільшенням обсягів даних та їх швидкою змінністю. Тобто обсяги даних збільшуються швидше, ніж здатність машин їх обробляти навіть за умов глобальної цифровізації. Звідси випливає низка наслідків.

- Методи та прийоми: деякі методи та прийоми, які добре зарекомендували себе в минулому, тепер потребують перегляду або заміни, тому що не масштабуються на сучасний обсяг даних.

- Обсяг даних: зі зростанням кількості джерел даних та збільшенням їх обсягу стає дуже важко зберігати, обробляти та аналізувати такі об'єми даних. Усі необхідні вихідні дані не можуть вже розміщуватися в оперативній пам'яті комп'ютерів. Застосування кластерів або багатоядерних процесорів стає потребою, а не розкішшю.

- Швидкість збільшення обсягів даних: дані можуть збільшуватися дуже швидко, що може ускладнити їх зберігання та обробку.

- Недостовірність даних: із збільшенням обсягів даних зростає і ймовірність того, що вони можуть бути недостовірними, неповними або некоректними. Це може виникати через помилки в процесі збору, обробки та аналізу даних.

- Структуризація даних: із збільшенням обсягів даних може зростати ймовірність того, що дані будуть неструктуровані. Це може зробити їх обробку та аналіз складнішими.

- Збереження даних: збільшення обсягів даних може потребувати значних витрат на їх збереження, зокрема, на обладнання, яке може забезпечувати достатню ємність для зберігання великих об'ємів даних.

- Недостатня кваліфікація персоналу: для ефективного аналізу та обробки даних потрібні спеціалісти з високою

кваліфікацією та знаннями новітніх технологій, але не завжди компанії можуть знайти відповідних спеціалістів.

### **3.1. *Поняття Big Data***

Big Data – це термін, що характеризує безліч наборів даних настільки об'ємних і складних, що унеможливує застосування наявних традиційних інструментів управління базами даних та додатків для їх обробки.

Проблему представляють збір, очищення, зберігання, пошук, доступ, передача, аналіз та візуалізація таких наборів як цілісних сутностей, а не локальних фрагментів.

Термін Big Data вперше з'явився у статті Кліффорд Лінч, редактор журналу Nature, 3 вересня 2008 року, який присвятив цілий спеціальний випуск журналів темі «які можуть означати для сучасної науки набори великих даних». До великих даних Лінч відніс будь-які масиви неоднорідних даних понад 150 Гб на добу, проте єдиного критерію досі немає.

Як визначальні характеристики для Big Data відзначають «три V»: обсяг (Volume, у сенсі величини фізичного обсягу), швидкість (Velocity, що означає в даному контексті швидкість приросту і необхідність високошвидкісної обробки та отримання результатів), різноманітність (Variety, в сенсі можливості одночасної обробки різних типів структурованих та напівструктурованих даних).

Наведені властивості даних є основними факторами, які зробили Big Data такою важливою для бізнесу та науки.

- Об'ємність (Volume) описує величину даних, які зберігаються та обробляються. В Big Data це можуть бути дуже великі обсяги даних, які зазвичай перевищують можливості традиційних баз даних та вимагають на



використання нових технологій зберігання та обробки даних, таких як Hadoop і Spark.

- Різноманітність (Variety) описує різноманітність джерел даних та форматів, в яких вони можуть знаходитись. Big Data можуть містити різноманітні дані, такі як тексти, зображення, відео, аудіо, сигнали датчиків, інтернет-трафік тощо. Різноманітність джерел даних, таких як соціальні мережі, датчики IoT та веб-сайти, вимагає розробки нових методів та алгоритмів обробки даних, таких як машинне навчання та аналіз тексту.

- Швидкість (Velocity) - швидке збільшення обсягу даних та необхідність обробки даних в режимі реального часу вимагає застосування нових методів та технологій обробки даних, таких як Apache Storm і Apache Kafka.

VVV у технологіях Big Data є ключовими складовими для розробки ефективних методів та інструментів для зберігання, обробки та аналізу великих обсягів даних.

У літературі з'явилися варіації з додатковими літерами V: (Veracity – достовірність, Viability – життєздатність, Value – цінність, Variability – мінливість та Visualization – візуалізація). У всіх випадках у цих ознаках підкреслюється, що визначальною характеристикою великих даних не лише їх фізичний обсяг, а й інші категорії, суттєві уявлення про складність завдання обробки та аналізу даних.

На цей час термін Big Data вже прижився і досяг піку використання. Час підтвердив справедливість виділення великих даних як окремого феномена. Сьогодні, згідно з дослідженнями агентства Gartner термін Big Data вже перевершив пік знаменитого гартнерівського Hype Cycle. Рис. 1 наведена статистика запитавши користувачів до системи Google за словосполучення "Big Data" (сервіс Google Trends, <https://trends.google.com/>).



Рисунок 1. Динаміка запитів «Big Data»

Розвиток напрямку Big Data пов'язаний із розвитком веб-простору та соціальних мереж, незважаючи на те, що великі обсяги даних притаманні таким галузям, як телекомунікаційна, енергетична, транспортна тощо.

Також розвиток напрямку Big Data безпосередньо пов'язаний з правом, оскільки збір, зберігання, обробка та використання даних потребують правової регуляції.

По-перше, використання великих обсягів даних може порушувати приватність людей, тому необхідно забезпечувати захист персональних даних та встановлювати механізми контролю за їх використанням. Наприклад, в Європейському Союзі існує Загальний регламент про захист персональних даних (GDPR), який встановлює правила збору, зберігання та використання персональних даних.

По-друге, використання даних може порушувати інтелектуальні права, зокрема авторські права, оскільки великі обсяги даних містять велику кількість інформації, яка може бути захищена авторським правом. Тому необхідно встановлювати механізми контролю за використанням такої інформації.

По-третє, використання даних може мати вплив на різні сфери життя людей, такі як здоров'я, освіта, зайнятість, економіка тощо, тому важливо забезпечувати прозорість та

відкритість використання даних та встановлювати механізми контролю за їх використанням.

Отже, використання технологій Big Data потребує розвитку відповідної правової бази, яка забезпечує захист прав людини, інтелектуальну власність, прозорість та відкритість використання даних.

Розвиток напрямку Big Data пов'язаний з правовими базами даних та їх характеристиками на різних рівнях. Наприклад, в області інтелектуальної власності, багато інформації може бути зібрано та оброблено великими компаніями, що збирають дані з різних джерел, таких як соціальні мережі, форуми, магазини тощо. Однак, збір та обробка таких даних може порушувати авторські права, права конфіденційності, а також створювати ризики щодо безпеки даних.

Правові бази даних також можуть бути використані для збору та аналізу великих обсягів даних з різних джерел, наприклад, з медичних записів або заявок на патенти. Однак, використання таких даних для досліджень може порушувати права на конфіденційність медичної інформації або інтелектуальної власності.

У контексті Big Data, також важливо забезпечити дотримання правил щодо захисту персональних даних, які збираються та обробляються в процесі збору даних.

Розвиток напряму Big Data також безпосередньо пов'язаний із парламентським контролем, оскільки об'ємні та складні дані, що генеруються в різних сферах життєдіяльності, можуть бути використані для підтримки роботи парламентських органів та забезпечення більш ефективного контролю за діяльністю влади та інших суб'єктів.

Big Data можуть допомогти парламентам збирати та аналізувати великі обсяги даних щодо діяльності владних структур, бізнесу, громадських організацій та громадськості.

Застосування концепції Big Data в правовій інформатиці може мати декілька переваг:

1. Обробка великих обсягів даних дозволяє здійснювати більш точний та швидкий аналіз правової інформації. Зокрема, застосування алгоритмів машинного навчання та штучного інтелекту може допомогти виявити зв'язки та закономірності в правових документах та судових рішеннях.
2. Застосування Big Data дозволяє збирати та аналізувати дані з різних джерел, що дозволяє отримувати більш повну картину про стан правопорядку, стан різних галузей права та ефективність діяльності правоохоронних органів та судів.
3. Використання Big Data може допомогти забезпечити більшу прозорість та відкритість у діяльності правоохоронних органів та судів. Наприклад, оприлюднення статистичної інформації про кількість затримань, проваджень та вироків може допомогти зменшити можливість корупції та порушень прав людини.
4. Застосування Big Data може допомогти вдосконалити процеси прийняття рішень в правовій сфері. Наприклад, на основі аналізу великих обсягів даних можна зробити прогноз щодо ефективності певної стратегії боротьби зі злочинністю або визначити тенденції у судовій практиці.
5. Застосування Big Data може допомогти знизити витрати на збір та аналіз правової інформації. Автоматизація процесів збору та аналізу даних

дозволяє зменшити кількість робочих годин та загальні знизити витрати.

***Питання для самоконтролю:***

- Які причини зростання обсягів даних, які потребують ефективної обробки та аналізу?
- Чому збільшення обсягів даних вимагає нових інструментів та технологій для ефективної обробки та аналізу?
- Що означає термін Big Data та хто вперше використав цей термін?
- Які критерії визначають великі дані, а саме "три V"?
- Які технології використовуються для зберігання та обробки великих обсягів даних, які перевищують можливості традиційних баз даних?
- Які проблеми можуть виникнути з використанням великих обсягів даних щодо приватності людей, і як їх можна уникнути?
- Які проблеми можуть виникнути з використанням великих обсягів даних щодо інтелектуальних прав, і як їх можна уникнути?
- Які саме джерела даних використовуються при зборі та аналізі правової інформації за допомогою концепції Big Data?
- Як оприлюднення статистичної інформації може допомогти забезпечити більшу прозорість та відкритість у діяльності правоохоронних органів та судів?

***3.2. Техніки у технології Big Data***

Техніки та технології Big Data є двома поняттями, які використовуються в контексті обробки великих обсягів даних. Основна різниця полягає в тому, що техніки використовуються для збору, обробки та аналізу даних, тоді

як технології забезпечують потужність та можливості для застосування технік.

Техніки Big Data охоплюють багато різних методів та інструментів, які дозволяють збирати, зберігати та обробляти великі обсяги даних швидко та ефективно. Ці техніки можуть використовуватись для аналізу даних в різних сферах, таких як медицина, маркетинг, фінанси та інші.

Технології Big Data, з іншого боку, дозволяють застосовувати техніки на великій швидкості та масштабі. Ці технології включають хмарні обчислення, віртуалізацію, архітектуру мікросервісів та інші. Вони дозволяють побудувати потужні, масштабовані та надійні інфраструктури для обробки великих обсягів даних.

У сфері права, Big Data техніки та технології можуть використовуватись для аналізу великих обсягів правових документів, таких як закони, судові рішення, звіти тощо. Вони можуть допомогти виявити тенденції та зв'язки між документами, розпізнати ключові терміни та поняття, та допомогти в прийнятті більш обґрунтованих рішень на основі аналізу даних.

### ***3.2.1. Техніки Big Data***

Техніки Big Data використовуються для збирання, зберігання, обробки та аналізу таких великих обсягів даних.

Основні методи зберігання та обробки даних в Big Data включають наступні:

1. Хмарні технології: це технології, які використовують віртуалізацію для зберігання та обробки даних на серверах віддалених центрів обробки даних. Такі технології включають Amazon Web Services, Microsoft Azure, Google Cloud Platform та ін.

2. Hadoop: це відкрите програмне забезпечення для зберігання та обробки великих обсягів даних. Hadoop використовується для розподіленої обробки даних на кластерах серверів.
3. NoSQL: це технології зберігання даних, які використовують нереляційні бази даних, які забезпечують більш гнучкий та масштабований підхід до зберігання даних. Такі технології включають MongoDB, Cassandra, Couchbase та ін.

Основні функціональні операції над даними, методи їх зберігання та обробки включають наступне:

1. Збір даних: збір та обробка даних з різних джерел в реальному часі, таких як датчики, соціальні медіа, веб-сайти тощо. Для цього використовуються технології, такі як Apache Kafka, Apache Flume, Logstash тощо.
2. Зберігання даних: зберігання даних великих обсягів забезпечується за допомогою різних технологій, таких як NoSQL-бази даних (MongoDB, Cassandra, HBase тощо) та розподілена файлова система Hadoop Distributed File System (HDFS).
3. Обробка даних: обробка даних зазвичай виконується за допомогою розподілених систем обробки даних, таких як Apache Hadoop та Apache Spark. Ці системи забезпечують можливість розподіленої обробки даних на багато вузлів, що дозволяє знизити час обробки великих обсягів даних.
4. Аналіз даних: аналіз даних виконується за допомогою різних методів, таких як машинне навчання, статистичний аналіз, аналіз тексту та інші методи аналізу даних. Для цього використовуються

різні інструменти та бібліотеки, такі як Apache Mahout, Apache Hadoop MapReduce, Apache Spark MLlib тощо.

5. Візуалізація даних: це процес відображення даних в графічному вигляді для полегшення розуміння та аналізу. Для цього використовуються інструменти, такі як Tableau, PowerBI, D3.js тощо.

Узагальнюючи вищезазначене можна виділити:

- консолідація даних;
- класифікація, кластеризація;
- машинне навчання;
- візуалізація.

### ***Консолідація даних***

Консолідація даних – це процес об'єднання даних з різних джерел в одне ціле. Цей процес включає збір, обробку, зведення та аналіз даних, щоб створити повну та точну картину про певну ситуацію або процес. Консолідація даних є важливим інструментом для бізнесу та управління, оскільки вона дозволяє отримати більш повне та об'єктивне уявлення про стан справ і приймати кращі рішення на основі цих даних. Цей процес може бути виконаний вручну або автоматично за допомогою спеціального програмного забезпечення.

Консолідація даних в праві може бути реалізована через різні підходи, наприклад, за допомогою аналітичних пісочниць (analytic sandbox) або озера даних (data lake).

Аналітична пісочниця є ізольованим середовищем для зберігання та аналізу даних, що містить підготовлені набори даних, інструменти для аналізу та візуалізації даних. Таке середовище дозволяє швидко отримувати результати аналізу



без впливу на вихідні дані. Аналітичні пісочниці використовуються в багатьох галузях, включаючи право, для підготовки та аналізу даних, які потрібні для прийняття рішень.

Озеро даних (data lake) - це архітектурний підхід, який забезпечує зберігання різноманітних даних у незмінному вигляді, включаючи структуровані, напівструктуровані та неструктуровані дані, великого обсягу, різної якості та походження. Це дозволяє збирати всі дані в одне місце, щоб мати можливість здійснювати подальшу обробку та аналіз даних за допомогою різних інструментів.

У праві консолідація даних може бути корисна для забезпечення більш детального та об'єктивного аналізу законодавства та правової практики, виявлення тенденцій та аномалій у різних галузях права, забезпечення ефективнішого прийняття рішень та більшої прозорості у процесі прийняття законодавчих актів.

Техніки консолідації даних включають:

1. Аналітика орієнтована на пакети даних (batch-oriented) - це техніка, яка полягає в обробці даних в пакетному режимі. Дані збираються, групуються в пакети та обробляються разом в заздалегідь встановлений час. Це дозволяє обробляти великі обсяги даних за короткий час, що робить цю техніку популярною для звітності та аналізу даних.
2. Аналітика в реальному часі (real-time oriented) - це техніка, яка забезпечує обробку даних в режимі реального часу, тобто негайну реакцію на події. Це дозволяє забезпечити миттєву реакцію на надходження нових даних та швидке прийняття рішень.
3. Гібридна аналітика (hybrid) - це техніка, яка поєднує переваги пакетної обробки даних та аналітики в

режимі реального часу. Вона дозволяє обробляти дані в реальному часі, одночасно зберігаючи їх для подальшого аналізу в пакетному режимі. Це забезпечує швидкий доступ до даних та можливість їх аналізу в будь-який момент часу.

Обрання техніки консолідації даних залежить від потреб користувачів та особливостей конкретної ситуації.

При пакетній аналітиці періодично виробляється вивантаження даних із різних джерел, дані аналізуються на наявність збійних фрагментів, шуму і їх фільтрація. За виконання аналітики реального часу дані виробляються джерелами безперервно і утворюють набір потоків даних. Аналіз цих потоків та своєчасне отримання результатів у заданому темпі вимагають забезпечити асинхронне отримання даних у вигляді деяких повідомлень та маршрутизувати ці повідомлення у потрібні процесингові вузли для обробки. Для гібридної аналітики зазвичай повідомлення даних повинні бути не тільки маршрутизовані на процесинг, а й інтегровані в аналітичну пісочницю для подальшої обробки за результатами накопичення даних за значні інтервали часу. Дані, отримані внаслідок консолідації, повинні відповідати певним критеріям якості. Якість даних - це критерій, що визначає повноту, точність, актуальність та можливість інтерпретації даних. Дані можуть бути високої та низької якості. Дані високої якості – це повні, точні, актуальні дані, що піддаються інтерпретації. Такі дані забезпечують отримання якісного результату: знань, які можуть підтримувати процес прийняття рішень. Основними процесами консолідації даних є так звані процеси ETL - Extraction-Transformation-Loading. Кожен з етапів відповідає певній функції:

- Extraction (вилучення) - процес отримання даних з джерел даних. Включає в себе вилучення даних зі

структурованих, напівструктурованих та неструктурованих джерел.

- Transformation (трансформація) - процес обробки даних, щоб зробити їх придатними для аналізу. Включає в себе перетворення, очищення, агрегацію, фільтрацію, об'єднання і трансформацію даних для подальшої обробки.
- Loading (завантаження) - процес завантаження оброблених даних до системи для збереження та аналізу. Включає в себе завантаження даних в бази даних, в файлові системи, в хмарні сервіси та інші системи для збереження та аналізу.

Після завершення процесів ETL, дані готові для аналізу та використання. Консолідовані дані можуть бути використані для побудови аналітичних звітів, прогнозування та прийняття рішень.

Додатки бізнес-аналітики до процесів ETL включалися дуже складні перетворення даних, такі як квантування, що дозволяє знизити обсяг оброблюваних даних, нормалізація – процес приведення реляційних таблиць до канонічного вигляду чи числових даних до єдиного масштабу, кодування даних – запровадження унікальних кодів для стиснення даних. У техніках великих даних зазвичай вважають, що необхідно працювати безпосередньо з брудними даними, оскільки нерідко саме характер збоїв може стати предметом аналізу, а стиснення даних є функцією власне аналітичних алгоритмів. Можливість зберігання даних у вихідному вигляді повинні надавати технічні засоби аналітичної системи. Якість великих даних нерідко важко оцінити методами формальних алгоритмів і тоді вдаються до візуалізації на ранньому етапі дослідження. Крім оцінки якості та вибору методу препроцесингу, візуалізація може

допомогти перейти до важливого етапу аналітики – вибору моделей, гіпотез для досягнення кінцевої мети – прийняття рішень.

### ***Візуалізація***

Додатки бізнес-аналітики до процесів ETL включалися дуже складні перетворення даних, такі як квантування, що дозволяє знизити обсяг оброблюваних даних, нормалізація – процес приведення реляційних таблиць до канонічного вигляду чи числових даних до єдиного масштабу, кодування даних – запровадження унікальних кодів для стиснення даних.

У техніках великих даних зазвичай вважають, що необхідно працювати безпосередньо з брудними даними, оскільки нерідко саме характер збоїв може стати предметом аналізу, а стиснення даних є функцією власне аналітичних алгоритмів.

Можливість зберігання даних у вихідному вигляді повинні надавати технічні засоби аналітичної системи. Якість великих даних нерідко важко оцінити методами формальних алгоритмів і тоді вдаються до візуалізації на ранньому етапі дослідження. Крім оцінки якості та вибору методу препроцесингу, візуалізація може допомогти перейти до важливого етапу аналітики – вибору моделей, гіпотез для досягнення кінцевої мети – прийняття рішень.

### ***Візуалізація даних з текстів***

Візуалізація текстів в контексті Big Data в області права є процесом перетворення великого обсягу текстової інформації на графічні зображення, які допомагають зрозуміти певні закономірності та тенденції в даних. Це може включати створення графіків, діаграм, хмар тегів та інших графічних зображень, що допомагають візуалізувати та аналізувати велику кількість текстуальної інформації.

Візуалізація текстів може бути корисною в багатьох аспектах правової діяльності, наприклад, при аналізі законодавства, судових рішень, договорів та інших юридичних документів. Вона дозволяє легше виявляти закономірності та тенденції в текстах, швидше знаходити потрібну інформацію та здійснювати більш об'єктивний аналіз.

Для візуалізації текстової інформації можна використовувати різні інструменти та програмні засоби, наприклад, Word Cloud Generator, Tableau, Gephi та інші. Важливим аспектом візуалізації є правильний вибір інструментів та налаштування параметрів, що дозволяють зробити графічні зображення зрозумілими та інформативними для аналізу.

Якщо дані є тексти природною мовою, то первинну допомогу в аналізі може надати візуалізація за допомогою розміченого тексту. Візуалізатор підраховує частоту згадок того чи іншого слова, і надає словам умовну вагу, яка залежить від цієї частоти. Слова різної ваги при візуалізації мають різну розмітку, а отже, різне подання на екрані. Одні слова виглядають більше за інші. Цей тип візуалізації допомагає досліднику швидко вхопити основні думки тексту.

### ***Візуалізація кластерів***

Однією з найчастіше використовуваних візуалізацій є візуалізація кластерів. Кластерами називають групи у чомусь схожих чи близьких за властивостями об'єктів. Алгоритми кластеризації, тобто. Розбиття безлічі об'єктів на групи, ми розглянемо нижче, а тут покажемо тільки як може бути візуалізована їхня робота. Більшість візуалізаторів підтримує алгоритми кластеризації та здатні розділяти дані на кластери.

Візуалізація кластерів в області права - це метод візуального аналізу даних, який дозволяє відобразити групу

подій, документів або законодавчих актів, які мають спільні характеристики та є схожими між собою.

Кластери можуть бути сформовані за допомогою алгоритмів кластеризації, які групують дані за певними ознаками, такими як ключові слова, тематика, географічне розташування, дата та інші параметри.

Після створення кластерів можна використовувати візуалізацію, щоб показати групування даних за допомогою графіків, діаграм та інших графічних зображень. Це дозволяє зробити аналіз даних більш зрозумілим та ефективним.

У правовій галузі візуалізація кластерів може бути корисною для аналізу судових рішень, статистичних даних, законодавчих актів та інших юридичних документів. Наприклад, можна візуалізувати кластери рішень судів за певною тематикою, щоб зрозуміти, які питання є найбільш актуальними в певній сфері права. Також можна візуалізувати кластери законодавчих актів, щоб виявити, які питання були найбільш обговорюваними та які є найбільш значущими для правової системи.

### ***Візуалізація асоціацій***

Візуалізація асоціацій демонструє частоту, з якою ті чи інші елементи з'являються разом у наборі даних, рахунок чого визначається структура організації даних (наприклад, може йтися у тому, які продукти часто продаються разом). Також можлива візуалізація інформації про силу асоціації даних.

Візуалізація асоціацій в області права є одним з методів аналізу даних з використанням технологій Big Data. Цей метод дозволяє виявляти зв'язки між різними елементами в правових текстах, такі як слова, терміни, поняття, норми тощо.

Для візуалізації асоціацій використовуються спеціальні інструменти, які дозволяють відобразити зв'язки між елементами у вигляді графіків, діаграм та інших візуальних форм. За допомогою таких інструментів можна виявляти найпоширеніші зв'язки між елементами в текстах та встановлювати певні закономірності в правовому просторі.

Наприклад, візуалізація асоціацій може допомогти виявити зв'язки між різними статтями законодавства, що дозволить більш ефективно аналізувати їх взаємозв'язок та виявляти можливі колізії чи прогалини в законодавстві. Також візуалізація асоціацій може бути корисною для аналізу судової практики та виявлення зв'язків між різними судовими рішеннями та їх інтерпретацією.

Загалом, візуалізація асоціацій є потужним інструментом аналізу даних в області права, який дозволяє здійснювати більш точний та комплексний аналіз правових текстів та встановлювати певні закономірності та зв'язки між елементами.

### ***Візуалізація гіпотез***

Візуалізація гіпотез дозволяє показувати виявлені закономірності, що підтверджують гіпотези, що висуваються. Подання інформації у різних візуалізаторах відрізняється. Наприклад, якщо рядки кругових 3D-діаграм відображають ознаки, використані класифікатором, кожна кругова діаграма відображає ймовірність того, що величина ознаки або діапазон значень підходять для класифікації.

Візуалізація гіпотез в області права полягає у візуальному поданні інформації про можливі зв'язки між різними елементами досліджуваного явища, випадку або проблеми, що досліджується. Такі зв'язки можуть бути встановлені на основі статистичного аналізу даних, аналізу текстів,

мережевого аналізу, машинного навчання та інших методів аналізу даних.

Візуалізація гіпотез зазвичай здійснюється за допомогою графічних зображень, таких як діаграми, графіки, сітки, мапи та інші візуальні інструменти. Наприклад, можна побудувати графік залежності між кількістю злочинів і кількістю поліцейських в різних районах міста, щоб проаналізувати можливі залежності між цими показниками. Також можна використовувати візуалізацію даних для ідентифікації проблемних ситуацій, які вимагають уваги з боку законодавців.

Одним з прикладів візуалізації гіпотез в області права є проект OpenLegalData, який розробляє нові методи та інструменти для візуалізації даних з різних правових джерел та показників правової діяльності. В рамках цього проекту використовуються інструменти машинного навчання, мережевого аналізу та інші методи для аналізу великих обсягів юридичних даних.

### ***Візуалізація дерев рішень***

Візуалізація дерев рішень дозволяє подати ієрархічно організовану інформацію у вигляді ландшафту та оглядати всю безліч даних або їх частину у вигляді вузлів та гілок. Ландшафт може бути як двовимірним, так і тривимірним.

Візуалізація дерев рішень є одним з методів візуалізації даних, що може застосовуватись в правовій сфері. Дерево рішень є графічним зображенням, яке демонструє послідовність виборів та відповідних результатів для різних сценаріїв прийняття рішень. В контексті права, дерева рішень можуть бути використані для відображення правових процесів, процедур та стандартів прийняття рішень, дозволяючи користувачам зрозуміти логіку та наслідки різних варіантів поведінки.



Для створення дерева рішень необхідно визначити множину можливих варіантів вибору та відповідних їм наслідків. Кожен варіант вибору представляє собою вузол у дереві, а наслідки - гілки, що виходять з вузла. Зазвичай дерева рішень створюються за допомогою спеціальних програм, які дозволяють збирати та організовувати дані.

Прикладами використання дерев рішень в правовій сфері можуть бути рішення суду або органів державної влади, що демонструють логіку та процедури прийняття рішень у відповідних випадках. Також дерева рішень можуть бути використані для розробки правових алгоритмів, які допомагають користувачам зрозуміти правила та процедури, які потрібно дотримуватися в конкретних випадках.

Візуалізація дерев рішень також може бути застосована в області парламентського контролю. Наприклад, можна використовувати дерева рішень для аналізу законопроектів та їх відповідності Конституції та іншим законодавчим актам. Дерева рішень можуть допомогти у формулюванні рекомендацій та прийнятті рішень щодо підтримки або відхилення законопроектів.

Також дерева рішень можуть використовуватись для аналізу процедур парламентського контролю, наприклад, для виявлення недоліків у законопроектах, що не проходять через комітети чи не беруться до розгляду на пленарному засіданні. Дерева рішень можуть допомогти візуалізувати ці процеси та знайти шляхи для їх оптимізації.

Один з прикладів застосування дерев рішень в парламентському контролі – це проект "Legislator", в рамках якого створено програмне забезпечення, призначене для аналізу законодавства та прийняття рішень у сфері права в США. Його розробниками є компанія Legal Robot, яка

спеціалізується на застосуванні штучного інтелекту та машинного навчання в правовій сфері.

Проект "Legislator" використовує технології обробки природньої мови та аналізу даних для автоматичного виявлення змін у законодавстві та надання корисної інформації про правову систему США. Він також надає можливість користувачам створювати свої власні проекти з аналізу законодавства та отримувати персоналізовані рекомендації щодо прийняття рішень у сфері права.

Основні переваги проекту "Legislator" полягають у здатності швидко та ефективно аналізувати великі обсяги законодавства та надавати корисну інформацію для прийняття рішень у сфері права. Він базується на використанні машинного навчання та аналізу даних для створення дерев рішень, які можуть бути використані для прогнозування результатів законодавчих процесів<sup>29</sup>.

Дерева рішень в "Legislator" використовуються для аналізу законодавства та передбачення його впливу на різні аспекти життя громадян. Вони допомагають встановити взаємозв'язки між різними аспектами законодавства та впливом цих аспектів на громадян. Також дерева рішень допомагають виявляти тенденції та прогнозувати результати законодавчих процесів на основі аналізу раніше прийнятих законів та їхнього впливу на суспільство.

Цей проект є прикладом використання візуалізації дерев рішень для аналізу законодавства та прийняття рішень у сфері права в парламентському контролі.

---

<sup>29</sup> Samuel Maireg Biresaw, Abhijit Umesh Saste. The Impacts of Artificial Intelligence on Research in the Legal Profession. *International Journal of Law and Society*. Vol. 5, No. 1, 2022, pp. 53-65. doi: 10.11648/j.ijls.20220501.17

## ***Класифікація***

Техніка класифікації одна із базових методик інтелектуального аналізу великих даних. Її нерідко використовують при побудові моделі аналітичних систем поряд із ще однією технікою – кластеризацією. Класифікація – це розподіл об'єктів (спостережень, подій) дослідження по заздалегідь відомим класам виходячи з подібності ознак. На відміну від класифікації кластеризація здійснює розподіл об'єктів (спостережень, подій) за невідомими заздалегідь класами. Класифікація проводиться у відповідності до принципів машинного навчання з учителем (Supervised Machine Learning). Для проведення класифікації за допомогою математичних методів необхідно мати формальний опис об'єкта, яким можна оперувати за допомогою математичного апарату класифікації. Кожен об'єкт (запис бази даних) повинен містити інформацію про деякі ознаки об'єкта.

Процес класифікації, як правило, зводиться до наступних кроків:

1. Набір вихідних даних (або вибірку даних) розбивають на дві множини: навчальна та тестова. Навчальна множина – множина, яка включає дані, що використовуються для конструювання моделі. Безліч містить вхідні та вихідні (цільові) значення прикладів. Вихідні значення призначені для навчання моделі. Тестова множина також містить вхідні та вихідні значення прикладів. Тут вихідні значення використовують для перевірки моделі.
2. Кожен об'єкт набору даних належить до одного визначеного класу. На цьому етапі використовується навчальна множина, на ньому відбувається конструювання моделі. Отримана модель є

класифікаційними правилами, деревом рішень або математичними формулами.

3. Оцінюється правильність моделі. Відомі значення тестової множини порівнюються з результатами використання отриманої моделі. Обчислюється рівень точності – відсоток правильно класифікованих об'єктів у тестовій множині.

Техніки класифікації як машинного навчання з учителем можуть застосовуватися в праві та парламентському контролі для різних завдань, наприклад:

1. Класифікація юридичних документів: великі обсяги юридичних документів, таких як законодавство, рішення судів, договори, можуть бути класифіковані за темою або категорією за допомогою моделей класифікації. Наприклад, можна створити модель, яка класифікує законодавчі акти за темою або за галуззю права.
2. Виявлення ризиків: машинне навчання може використовуватися для класифікації правових питань, що стосуються ризиків, наприклад, класифікація ризиків для страхових компаній або класифікація питань, які можуть мати негативний вплив на фінансові показники підприємства.
3. Класифікація виборчих даних: класифікація виборчих даних може допомогти в політичному аналізі та в парламентському контролі, наприклад, для класифікації депутатів за політичною партією або за голосуванням щодо певної теми.
4. Класифікація ризиків корупції: машинне навчання може використовуватися для класифікації підозрілих транзакцій або дій, які можуть свідчити про корупцію.

Це може допомогти виявити можливі порушення і прийняти відповідні заходи.

5. Класифікація порушень прав людини: машинне навчання може бути використане для класифікації порушень прав людини за різними параметрами, наприклад, за країною, типом порушення, суб'єктом порушення тощо.

### ***Кластерний аналіз***

Техніка кластеризації є підходом до класифікації даних у разі, коли заздалегідь невідомо, якого класу має бути віднесено будь-який з наявних об'єктів. Кластеризація здійснюється автоматичним перебуванням груп, куди мають бути розбиті аналізовані об'єкти. Такий процес можна розглядати як машинне навчання без вчителя (Unsupervised Machine Learning).

Техніки кластерного аналізу в машинному навчанні без учителя можуть бути застосовані в праві та парламентському контролі для розподілу об'єктів (наприклад, законів або документів) на групи або кластери в залежності від їх характеристик або властивостей. Наприклад, застосування кластерного аналізу може допомогти у визначенні груп законопроектів за їх тематикою або за ступенем важливості для парламенту.

Для прикладу, можна провести кластерний аналіз нормативно-правових актів з метою виявлення схожих груп законів або документів. Для цього можна використовувати алгоритми кластерного аналізу, такі як ієрархічний кластерний аналіз або кластерний аналіз на основі методу k-середніх. Результатом такого аналізу може бути групування законів за їх тематикою або ступенем важливості для парламенту, що може допомогти в процесі прийняття рішень.

Також кластерний аналіз може використовуватись для аналізу розподілу виборців за різними параметрами (наприклад, віком, статтю, освітою тощо). Це може допомогти парламентаріям зрозуміти склад виборців у певному регіоні та визначити найбільш важливі питання для цієї групи людей.

Можливі приклади застосування технік кластерного аналізу в парламентському контролі включають:

1. Кластеризація законопроектів за темами: кластерний аналіз може допомогти згрупувати законопроекти за їх темою або напрямком, що допоможе політикам і аналітикам в легшому розумінні їх змісту і сприятиме прийняттю рішень.
2. Сегментація виборців за інтересами: застосування кластерного аналізу може допомогти сегментувати виборців за їх інтересами та поглядами, що дозволить політикам створювати більш ефективні стратегії комунікації з виборцями.
3. Кластеризація законодавчих дій: кластерний аналіз може допомогти згрупувати законодавчі дії за їх характеристиками, такими як тема, вид законодавчої ініціативи, стан законодавчої процедури тощо. Це дозволить політикам та аналітикам зрозуміти, які законодавчі дії найбільш важливі для дальшого аналізу.
4. Аналіз законодавчого впливу: кластерний аналіз може допомогти визначити, які законодавчі ініціативи мають найбільший вплив на різні сфери діяльності, такі як економіка, наука, соціальні програми тощо.

### ***Машинне навчання***

Термін «машинне навчання», найімовірніше, зустрічався вам неодноразово. Хоча його нерідко використовують як синонім штучного інтелекту, насправді машинне навчання – це один із його елементів. При цьому обидва поняття народилися у Массачусетському технологічному інституті наприкінці 1950-х років.

Машинне навчання (machine learning, ML) - клас методів штучного інтелекту, характерною рисою яких є не пряме розв'язання задачі, а навчання в процесі застосування рішень безлічі подібних задач. Для побудови таких методів використовуються засоби математичної статистики, чисельних методів, методів оптимізації, теорії ймовірностей, теорії графів, різних технік роботи з даними в цифровій формі.

Розрізняють два типи машинного навчання:

1. Навчання за прецедентами, або індуктивне навчання, засноване на виявленні емпіричних закономірностей даних.
2. Дедуктивне навчання передбачає формалізацію знань експертів та його перенесення на комп'ютер як бази знань. Дедуктивне навчання прийнято відносити до галузі експертних систем, тому терміни машинне навчання та навчання за прецедентами можна вважати синонімами.

Багато методів індуктивного навчання розроблялися як альтернатива класичним статистичним підходам. Багато методів тісно пов'язані з вилученням інформації (Information Extraction, Information Retrieval), інтелектуальним аналізом даних (Data Mining).

На відміну від традиційного програмного забезпечення, яке чудово справляється з виконанням інструкцій, але не здатне до імпровізації, системи машинного навчання, по суті,

програмують самі себе, самостійно розробляючи інструкції шляхом узагальнення відомих відомостей.

Класичний приклад машинного навчання – це розпізнавання образів або комп'ютерний зір - це технологія машинного навчання, яка може бути корисною в правовій практиці для автоматичного аналізу великих обсягів зображень, зокрема фотографій та відео. Нижче наведено кілька прикладів застосування розпізнавання образів в правовій практиці:

1. Розпізнавання образів для автоматичного розпізнавання номерних знаків автомобілів – ця технологія може бути корисною для правоохоронних органів при розслідуванні кримінальних справ, зокрема злочинів, пов'язаних зі злочинними автомобілями або протиправним перетином кордону.
2. Розпізнавання образів для автоматичної оцінки пошкоджень в аварійних ситуаціях – ця технологія може бути корисною для страхових компаній та правоохоронних органів при вирішенні питань про відшкодування збитків, пов'язаних з аварійними ситуаціями на дорозі.
3. Розпізнавання образів для визначення розташування та ідентифікації осіб на відеозаписах - ця технологія може бути корисною для правоохоронних органів при розслідуванні кримінальних справ, зокрема злочинів, пов'язаних зі злочинними діями на вулиці або в громадських місцях.
4. Розпізнавання образів для автоматичної ідентифікації злочинців на фотографіях - ця технологія може бути корисною для правоохоронних органів при розслідуванні кримінальних справ та пошуку злочинців, які знаходяться у розшуку.



### *Навчання з учителем і без нього*

Машинне навчання з учителем (supervised machine learning) - це підхід до машинного навчання, при якому алгоритм навчається за допомогою позначених прикладів. Іншими словами, для машинного навчання з учителем потрібно мати набір даних, в якому для кожного вхідного прикладу визначений вихідний результат або позначка, яку потрібно передбачити для нових прикладів.

Мета машинного навчання з учителем полягає в тому, щоб побудувати модель, яка може здійснювати передбачення для нових прикладів. Для цього модель навчається на наборі даних, де для кожного вхідного прикладу відомий його вихідний результат. Модель навчається шляхом визначення кореляції між вхідними даними та їх вихідними результатами, з використанням різних статистичних та математичних методів.

Після навчання модель може бути використана для передбачення вихідних результатів для нових вхідних прикладів, що не були використані під час навчання. Наприклад, якщо ми маємо набір даних з характеристиками клієнтів та їх покупками, то за допомогою машинного навчання з учителем ми можемо побудувати модель, яка буде передбачувати, які товари будуть куплені клієнтом в майбутньому.

Типові завдання навчання з учителем – класифікація та прогнозування (або регресійний аналіз). Деякі з найпоширеніших методів машинного навчання з учителем включають:

1. Логістична регресія (Logistic Regression): використовується для класифікації даних на два або більше класів.

2. Дерева рішень (Decision Trees): будується дерево, яке допомагає зрозуміти взаємозв'язок між вхідними даними та вихідним результатом.
3. Випадковий ліс (Random Forest): використовується для класифікації або регресії шляхом комбінування результатів декількох дерев рішень.
4. Наївний Байєс (Naive Bayes): використовується для класифікації даних на основі ймовірності.
5. Метод опорних векторів (Support Vector Machines): використовується для класифікації або регресії шляхом знаходження оптимальної лінії або площини, яка розділяє класи.
6. Нейронні мережі (Neural Networks): використовується для класифікації або регресії шляхом створення архітектури нейронної мережі та навчання ваги між нейронами.
7. Градієнтний бустинг (Gradient Boosting): використовується для покращення результатів моделі шляхом послідовного додавання слабких моделей до сильної моделі.

Ці методи можуть бути використані для різноманітних задач, таких як передбачення вартості нерухомості, категоризації зображень, визначення емоцій на обличчі, передбачення ризику захворювання та багато інших.

Деякі з методів машинного навчання з учителем, які можуть застосовуватися у правовій інформатиці та парламентському контролі, включають:

1. Класифікація текстів: цей метод може бути використаний для автоматичного класифікування текстових документів за різними категоріями, такими як рішення суду, законопроекти, резолюції та інші.

2. Аналіз настрою текстів: цей метод може бути використаний для автоматичного аналізу настрою текстових документів, таких як заяви, новини та соціальні медіа, щоб визначити тон тексту.
3. Автоматичне визначення мови: цей метод може бути використаний для автоматичного визначення мови тексту, що допомагає відокремити документи на різних мовах.
4. Передбачення рішень суду: цей метод може бути використаний для передбачення рішень суду на основі аналізу попередніх рішень суду та факторів, що впливають на них.
5. Системи рекомендацій: цей метод може бути використаний для автоматичного рекомендування законопроектів або резолюцій.

Машинне навчання без учителя – це підхід до машинного навчання, в якому модель аналізує дані без заздалегідь визначених міток або категорій. У цьому підході модель самостійно знаходить корисні залежності та патерни в навчальних даних, без допомоги вчителя, і використовує ці знання для подальшого аналізу даних або прийняття рішень.

Основні методи машинного навчання без учителя включають кластеризацію, зменшення розмірності та виявлення аномалій.

У кластеризації модель намагається розділити навчальні дані на групи (кластери), які мають схожі характеристики. Зменшення розмірності зменшує кількість змінних, які враховує модель, зберігаючи при цьому значущі залежності між змінними. Виявлення аномалій дозволяє моделі ідентифікувати незвичайні або відхилені від стандартних шаблонів дані.

Машинне навчання без учителя широко використовується у багатьох галузях, включаючи бізнес-аналітику, медицину, соціальні науки та інші.

Системи навчання без вчителя можуть виявляти закономірності у великих обсягах даних набагато швидше, ніж люди. Саме тому банки використовують їх для виявлення шахрайських операцій, маркетологи – для ідентифікації клієнтів зі схожими атрибутами, а ПЗ безпеки – для розпізнавання шкідливої активності у мережі.

Нижче наведено декілька прикладів завдань машинного навчання без учителя в правовій інформатиці та парламентському контролі:

1. Кластеризація текстів законодавства: Задача полягає в тому, щоб згрупувати закони та інші нормативні акти відповідно до схожості їх змісту. Це дозволить здійснювати більш ефективний пошук та аналіз нормативної бази.
2. Виявлення тематичних зв'язків між законопроектами: Задача полягає в тому, щоб автоматично визначати тематичні зв'язки між законопроектами на основі аналізу текстів законопроектів. Це допоможе підвищити ефективність парламентської діяльності та забезпечити більш якісний парламентський контроль.
3. Виявлення незвичайної активності у соціальних мережах: Задача полягає в тому, щоб автоматично виявляти незвичайну активність у соціальних мережах, що може бути пов'язано з поширенням фейкових новин або зловживанням правами на вираз своєї думки. Це допоможе забезпечити більш ефективний контроль за дотриманням прав та свобод людей у мережі.

4. Виявлення аномалій у законодавчих актах: Задача полягає в тому, щоб ідентифікувати законодавчі акти, які відрізняються від звичайних шаблонів законодавства. Це може допомогти виявити нетипові або підозрілі положення в законодавчих актах та надати рекомендації з приводу їх подальшої обробки.
5. Зменшення розмірності законодавчих даних: Задача полягає в тому, щоб зменшити кількість змінних, які враховуються при аналізі законодавчих актів, зберігаючи при цьому значущі залежності між ними. Це може допомогти зменшити складність аналізу законодавчих даних та забезпечити більш ефективну обробку даних.
6. Кластеризація депутатських груп: Задача полягає в тому, щоб класифікувати депутатів за їх голосуванням, тематикою інтересів та іншими характеристиками, групуючи їх у кластери. Це може допомогти розуміти залежності між різними депутатськими групами та допомогти у формуванні різних парламентських коаліцій та альянсів.

### ***Способи машинного навчання***

Розділ машинного навчання сформувався в результаті розвитку інформатики та штучного інтелекту.

З одного боку, розвиток обчислювальної техніки та збільшення обсягів зберігання та обробки даних зробили можливим застосування нових методів обробки даних, які згодом згрупувалися в машинне навчання.

З іншого боку, машинне навчання є частиною штучного інтелекту, тому що воно займається розробкою алгоритмів та програм, які навчають комп'ютери розуміти інформацію та приймати рішення на основі даних.

Таким чином, розвиток інформаційних технологій та штучного інтелекту сприяв створенню машинного навчання, яке здатне вирішувати різноманітні завдання на основі аналізу даних та здобуття знань.

Машинне навчання, штучний інтелект і математична статистика пов'язані між собою і взаємодіють у процесі розробки та використання алгоритмів та моделей.

Математична статистика надає математичний апарат для опису та аналізу даних, який використовується в машинному навчанні. Математичні методи статистики, такі як лінійна регресія, аналіз дисперсії та кластерний аналіз, використовуються для побудови моделей та прогнозування на основі даних.

Штучний інтелект, зокрема, нейронні мережі, використовується для створення засобів машинного навчання. Нейронні мережі забезпечують комп'ютеру здатність самостійно навчатися та розв'язувати завдання, що раніше потребували людського втручання.

Таким чином, машинне навчання поєднує математичну статистику та штучний інтелект, щоб створити інтелектуальні системи, які можуть самостійно вирішувати різноманітні завдання.

Методи машинного навчання можна класифікувати в залежності від способу навчання на наступні типи:

1. Навчання з вчителем (Supervised learning) - це метод машинного навчання, при якому існує набір вхідних даних та відповідний набір вихідних даних, з яких модель може навчитися передбачати відповідність між ними. Приклади задач навчання з вчителем: класифікація, регресія, сегментація зображень тощо.

2. Навчання без вчителя (Unsupervised learning) - це метод машинного навчання, при якому немає вихідних даних для навчання моделі, і вона повинна знаходити корисну інформацію самостійно шляхом аналізу вхідних даних. Приклади задач навчання без вчителя: кластеризація, зменшення розмірності, пошук аномалій.
3. Навчання з підсиленням (Reinforcement learning) - це метод машинного навчання, при якому модель навчається приймати рішення на основі взаємодії з навколишнім середовищем та отриманням винагороди або штрафу за свої дії. Приклади задач навчання з підсиленням: гра в шахи, навчання робота здійснювати рухи, щоб досягти мети.
4. Півнавчання (Semi-supervised learning) - це метод машинного навчання, при якому частина вхідних даних містить вихідні дані, а інша частина - ні. Модель навчається на доступних даних, а потім застосовує отримані знання для прогнозування результатів для недоступних даних.
5. Активне навчання (Active learning) - це метод машинного навчання, при якому модель вивчається на основі ітеративної взаємодії з людьми-експертами.

Алгоритми машинного навчання потребують даних, якнайбільшої кількості даних з якомога ширшого набору джерел. Чим більше вони "живляться" цими даними, тим "розумнішими" стають і тим більший їхній потенціал при прийнятті рішень. І хмари дають ці великі дані.

Big Data обіцяють нам знайти багато цінного в процесі цифрової трансформації, а хмарні обчислення пропонують будівельні блоки для цього процесу. Машинне навчання, своєю чергою, стало першим по-справжньому промисловим

інструментом для масштабного освоєння цих нових цінностей. Привабливість машинного навчання у цьому, що його використання практично безмежні. Воно може застосовуватися скрізь, де важливий швидкий аналіз даних, і зробити революційний ефект там, де важливо виявляти тенденції або аномалії в великих наборах даних — від клінічних досліджень до сфери безпеки та контролю за дотриманням стандартів. Машинне навчання має деякі привабливі переваги для застосування в сфері парламентського контролю, серед яких можна виділити наступні:

1. Обробка великого обсягу даних: парламентський контроль повинен мати можливість вивчати великі обсяги даних, що можуть стосуватися різних аспектів життя. Машинне навчання дозволяє швидко та ефективно обробляти великий обсяг даних.
2. Автоматизація процесу аналізу: машинне навчання може допомогти автоматизувати процес аналізу даних та виявлення невідповідностей у законодавстві та роботі державних установ, що сприяє більш ефективному парламентському контролю.
3. Покращення точності аналізу: машинне навчання може допомогти покращити точність аналізу даних та виявлення невідповідностей у законодавстві та діяльності державних установ, що може значно зменшити кількість помилок та недоліків, що можуть стати причиною негативних наслідків.
4. Пошук взаємозв'язків: машинне навчання може допомогти знайти взаємозв'язки між різними показниками та факторами, що можуть впливати на ефективність парламентського контролю, та допомогти прийняти раціональні рішення.



5. Мінімізація людського фактору: машинне навчання дозволяє мінімізувати вплив людського фактору на результати аналізу та прийняття рішень, що може зменшити ймовірність виникнення помилок та підвищити об'єктивність процесу.

Застосування машинного навчання в сфері парламентського контролю може допомогти покращити ефективність та якість роботи парламенту, зменшити кількість помилок та недоліків, а також забезпечити більш об'єктивні результати аналізу та прийняття рішень.

### ***Обмеження машинного навчання***

Кожна система машинного навчання створює власну схему зв'язків, являючи собою щось на зразок чорного ящика. Не можливо шляхом інженерного аналізу з'ясувати, як саме виконується класифікація, але це найчастіше не має значення, головне, щоб працювало.

Незважаючи на потенційні переваги, застосування машинного навчання в сфері парламентського контролю також має свої обмеження:

1. Недостатня кількість даних: для успішного застосування машинного навчання потрібна велика кількість якісних даних. Однак в деяких сферах парламентського контролю, наприклад, в сфері економіки або соціальних показників, може бути складно зібрати достатню кількість даних для успішної роботи моделей машинного навчання.
2. Складність обробки даних: деякі види даних можуть бути складні для обробки моделями машинного навчання, особливо якщо ці дані містять багато шуму або неоднорідні.

3. Ризик упередженості: моделі машинного навчання можуть бути упереджені через відсутність репрезентативної вибірки даних або неправильної специфікації моделі. Це може призвести до некоректних або небажаних результатів.
4. Обмеження щодо пояснення результатів: деякі методи машинного навчання, особливо нейронні мережі, можуть бути складні для інтерпретації, тобто складно пояснити, як саме модель прийшла до певного результату. Це може створювати проблеми з оцінкою правильності та надійності результатів.
5. Етичні питання: застосування машинного навчання в сфері парламентського контролю може виникати етичні питання, пов'язані з захистом особистої інформації та конфіденційності. Наприклад, може виникнути проблема використання даних громадян без їх згоди або порушення їх прав на конфіденційність.

Можна вважати, що система машинного навчання корисна лише настільки, наскільки точні навчальні дані: якщо подати їй на вхід «сміття», то результат буде відповідним. При неправильному навчанні або занадто малому розмірі навчальної вибірки алгоритм може видавати неправильні результати.

Ми розглянемо базові технології та інструменти, які сьогодні набули найбільшого поширення у відомих проектах. Цей список не вичерпує всіх вже апробованих технологій і тим більше що знаходяться в розробці, проте він дозволяє отримати досить цілісне уявлення про те, чим користуються сьогодні дослідники даних і якими інструментами необхідно володіти, щоб розгорнути проект з використанням великих даних.

### **3.2.2. Технології Big Data**

Більшість найбільших постачальників інформаційних технологій для організацій у своїх ділових стратегіях використовують поняття про великі дані, у тому числі IBM, Oracle, Microsoft, Hewlett-Packard, EMC, а основні аналітики ринку інформаційних технологій присвячують концепції виділених досліджень.

Нижче розглядаються базові, найбільш поширені технології та інструменти, якими необхідно володіти, щоб вести сьогодні проекти в галузі правової інформатики з використанням Big Data, зокрема, при створенні сучасних правових аналітичних систем (ПАС) з елементами контент-моніторингу інформації з Інтернету. Будуть представлені основні технологічні рішення, які використовуються в таких компонентах ПАС, як:

- сервери збору інформації, тобто інформаційні проксі-сервери, що забезпечують збирання інформації, розташованої у зовнішніх центрах обробки даних (ці сервери, з одного боку, призначені для надання надійних послуг користувачам корпоративних мереж, а з іншого боку можуть забезпечити обмін даними з подібними зовнішніми серверами). Вони можуть отримувати дані за сценаріями, визначеними адміністратором безпосередньо з інтернет-ресурсів, або через посередників-агрегаторів інформації. Такі сервери необхідні, крім технологічних причин, ще й у святі з правовими обмеженнями доступу до інформації у різних державах;
- сервер аналітики (сервер здійснює аналітичну обробку інформації та пошук інформації. За допомогою сервера підтримуються бази даних історичної (архівної) інформації. Аналітична обробка інформації включає:

вилучення понять; підтримка геоінформації; визначення тональності (сентимент-аналіз); агрегування інформації; аналіз динаміка інформації за певними напрямками, прогнозування динаміки інформаційних потоків, аналіз властивостей джерел інформації тощо);

- інтерфейсний сервер (сервер, з якого кінцеві користувачі можуть отримати доступ через веб-браузер, різні агрегатори або через API-додатки до системних ресурсів).

За останнє десятиліття з'явилися різні системи зберігання та обробки великих масивів даних. Серед них можна виділити проекти екосистеми Hadoop, деякі бази даних (БД) NoSQL, а також пошукові та аналітичні системи на кшталт Elasticsearch.

Сучасний технологічний комплекс правової інформації може базуватися на таких технологічних компонентах як модулі збору інформації в мережі; інформаційно-пошукова система (наприклад, Elasticsearch); засоби узагальненого доступу до даних типу Kibana; графові системи управління базами даних Neo4j; засобів візуалізації результатів Такий комплекс зможе забезпечувати реалізацію таких функцій, як:

- формування баз даних за певними інформаційними ресурсами;
- ведення баз даних;
- повнотекстовий пошук; аналіз текстових документів;
- аналіз та візуалізацію даних, у тому числі дослідження динаміки тематичних інформаційних потоків;
- прогнозування розвитку подій на основі аналізу динаміки публікацій тощо.

Нижче розглядаються:

- можливості технологій аналізу даних та застосування наукових методів, у тому числі методів інтелектуального аналізу даних, у Big Data;
- особливості архітектурних рішень при створенні та розгортанні систем обробки Big Data, а також вибір технології зберігання та обробки Big Data, використання сучасних високопродуктивних систем зберігання та обробки великих даних;
- основні технології та інструменти роботи з великими дачами, такі як Elastic Stack, Elasticsearch, Kibana, Neo4j, MongoDB;
- компоненти програмного забезпечення, необхідні для роботи з правовою інформацією у розподілених інформаційних системах Big Data.

Як провідний сучасний засіб агрегації великих обсягів правових документів, їх пошуку, обробки візуалізації розглядається екосистема компонентів Elastic Stack<sup>30</sup>.

Elasticsearch<sup>31</sup> – це інформаційно-пошукова система (ядро Elastic Stack), яка дозволяє здійснювати обробку неструктурованих даних, інформаційний пошук, аналіз даних, забезпечує підтримку користувацьких бібліотек та REST API, легке управління та масштабування. Утиліта Kibana – це вікно в Elastic Stack, засіб маніпуляції, аналізу та візуалізації інформації, що реалізує такі види відображення даних з Elasticsearch, як гістограми, карти, лінійні графіки, часові ряди.

---

<sup>30</sup> Learning Elastic Stack 6.0: A beginner's guide to distributed search, analytics, and visualization using Elasticsearch, Logstash and Kibana (Source Code). Pranav Shukla, Sharath Kumar M. N. Packt Publishing, 2017. – 434 pp.

<sup>31</sup> Elasticsearch: The Definitive Guide / Clinton Gormley and Zachary Tong. – O'Reilly Media, Inc., 2015. – 719 pp.

На окрему увагу в правовій інформації заслуговують засоби аналізу мережевих структур, графових СУБД. Нижче розглядаються можливості двох основних систем – програм аналізу та візуалізації графів Gephi та графова система управління базами даних Neo4j<sup>32</sup>. Серед особливостей програми Gephi вивчаються інтерфейс користувача, можливості компонування графів, фільтрація, дослідження даних, візуалізація, підтримка графічних форматів даних. Графова СУБД Neo4j забезпечує збереження та обробку мережевих даних великих обсягів, включає декларативну мову запитів до графових структур Cypher.

Технології великих даних повинні забезпечувати рішеннями та інструментами, що дозволяють реалізовувати описані вище техніки на значних обсягах різнорідних даних із необхідною швидкістю. Досягається це високою паралелізацією обчислень та розподіленням зберігання даних. Незважаючи на потребу значної обчислювальної потужності та пам'яті, як правило, розгортання програмних продуктів великих даних проводиться на кластерах із комп'ютерів середнього або навіть низького класу (commodity computers). Це дозволяє масштабувати системи великих даних без залучення суттєвих витрат. Останнім часом для розгортання систем великих даних все ширше використовуються хмарні послуги (cloud computing services). У разі імплементації системи у хмарі вузли обчислювального кластера реалізуються на віртуальних машинах хмарної інфраструктури та гнучко адаптуються до завдання, знижуючи витрати на використання. Це є додатковим фактором, що залучає багатьох розробників будувати системи великих даних на хмарних платформах.

---

<sup>32</sup> Graph Databases. Ian Robinson, Jim Webber, Emil Eifrem. – O'Reilly Media, 2013. – 201 pp.

Найбільш популярною технологією великих даних, яка вважається де-факто стандартом для побудови систем аналітики, що працюють у пакетному режимі, є сукупність рішень та програмних бібліотек, об'єднаних під назвою Hadoop. Якщо великі дані надходять у вигляді високошвидкісних потоків і реагування системи має відбуватися з малою затримкою, замість пакетної аналітики застосовується аналітика реального часу.

### ***Apache Hadoop***

За назвою Hadoop спільнота Apache просуває технологію, засновану на використанні спеціальної інфраструктури для паралельної обробки великих обсягів даних. Hadoop забезпечує середовище для функціонального програмування завдань, автоматичного розпаралелювання робіт, усунення обчислювального навантаження до даних. Hadoop створив Дуг Каттінг – творець Apache Lucene, широко використовуваної бібліотеки текстового пошуку. Hadoop походить від Apache Nutch - системи веб-пошуку з відкритим кодом, яка сама по собі була частиною проекту Lucene. Історія Hadoop безпосередньо пов'язана з розробкою Google File System (2003) і потім реалізацією технології MapReduce (2004). На основі цих компонентів у 2005 році з'явився додаток пошуку інформації Apache Nutch, який на наступний рік дав дорогу проекту Apache Hadoop.

### ***Стек Elastic***

Elastic Stack – велика екосистема компонентів, які служать для пошуку та обробки даних. Основні компоненти Elastic Stack - це Kibana, Logstash, Beats, X-Pack та Elasticsearch. Ядром Elastic Stack є пошукова система Elasticsearch, яка надає можливості для зберігання, пошуку та обробки даних. Утиліта Kibana, яку також називають вікном в Elastic Stack, є відмінним засобом візуалізації та користувацьким

інтерфейсом для Elastic Stack. Компоненти Logstash та Beats дозволяють передавати дані до Elastic Stack. X-Pack надає потужний функціонал: можна налаштовувати моніторинг, додавати різні повідомлення, встановлювати параметри безпеки для підготовки вашої системи до експлуатації.

### *Neo4j*

Neo4j – графова система управління базами даних з відкритим вихідним кодом, мовою Java, з підтримкою транзакції (ACID – Atomicity, Consistency, Isolation, Durability) — це набір властивостей, що гарантують надійну роботу транзакцій бази даних: атомарність, узгодженість, ізолюваність, довговічність). Станом на 2023 рік Neo4j вважається найпоширенішою графовою СУБД. Розробник – американська компанія Neo Technology, розробка ведеться вже 20 років.

Дані зберігає у своєму форматі, спеціалізовано пристосованому для подання графової інформації, такий підхід у порівнянні з моделюванням графової бази даних засобами реляційної СУБД дозволяє застосовувати додаткову оптимізацію у разі даних із складнішою структурою. Також стверджується про наявність спеціальних оптимізацій для SSD-накопичувачів, при цьому для обробки графа не потрібне його поміщення цілком в оперативну пам'ять обчислювального вузла, таким чином, можлива обробка досить великих графів.

Основні сфери застосування: соціальні мережі, системи надання рекомендацій, виявлення шахрайства, картографічні системи.

### *NoSQL СУБД*

Під NoSQL (від англ. Not Only SQL – не тільки SQL) розуміється широкий клас сучасних різнорідних систем управління базами даних, які істотно відрізняються від



традиційних реляційних СУБД з доступом до даних засобами мови SQL. Такі СУБД застосовуються до систем, у яких робиться спроба вирішити проблеми масштабованості та доступності за рахунок повної або часткової відмови від вимог атомарності та узгодженості даних<sup>33</sup>.

Традиційні СУБД орієнтуються на вимоги ACID до транзакційної системи: Atomicity (атомарність), Consistency (узгодженість), isolation (ізолюваність), durability (довговічність), тоді як у NoSQL замість ACID може розглядатися набір властивостей BASE:

- базова доступність (Basic Availability) – кожен запит гарантовано завершується (успішно чи безуспішно).
- гнучкий стан (Soft state) — стан системи може змінюватися з часом, навіть уведення нових даних, задля досягнення узгодження даних.
- узгодженість у кінцевому рахунку (Eventual consistency) – дані можуть бути деякий час неузгоджені, але приходять до погодження через деякий час.

Термін «BASE» було запропоновано Еріком Брюером, автором теореми CAP.

Теорема CAP стверджує, що можна створити розподілену систему, яка буде узгоджена (consistent) (тобто операції запису атомарні, і всі наступні операції читання бачать нове значення), доступною (available) (база даних повертає значення, поки працює хоча б один сервер) та стійкою до втрати зв'язності (partition tolerant) (система продовжує функціонувати, навіть якщо зв'язки між серверами тимчасово

---

<sup>33</sup> Seven Databases in Seven Weeks: A Guide to Modern Databases and the NoSQL Movement. Luc Perkins, Eric Redmond, Jim Wilson. – Pragmatic Bookshelf, 2018. – 360 pp.

відсутні), але одночасно можна гарантувати лише дві з цих трьох властивостей.

Таким чином, можна створити розподілену систему, яка буде узгодженою та стійкою до втрати зв'язності, систему, яка буде доступною та стійкою до втрати зв'язності, або систему, яка буде узгодженою та доступною (але при цьому не буде стійкою до втрати зв'язності, тобто, по суті, не буде розподіленою). Але неможливо створити розподілену базу даних, яка була б одночасно узгодженою, доступною та стійкою до втрати зв'язності.

Теорема CAP істотна під час проектування баз даних, оскільки у своїй слід вирішити, чим готові пожертвувати. Яку б базу даних ви не обрали, вона не зможе гарантувати доступність або узгодженість. Стійкість до втрати зв'язності – суто архітектурне рішення (від нього залежить, чи система взагалі буде розподіленою). Важливо розуміти сенс теореми CAP, щоб реалістично оцінювати можливості. Компроміси, прийняті у різних описаних у книзі базах даних, спираються саме на цю теорему.

Зрозуміло, що системи на основі BASE не можуть використовуватися в будь-яких додатках: для функціонування біржових та банківських систем використання транзакцій є необхідністю. У той же час властивості ACID, хоч би якими бажаними вони були, практично неможливо забезпечити в системах з багатомільйонною веб-аудиторією, на зразок amazon.com. Проектувальники NoSQL-систем жертвують узгодженістю даних для досягнення двох інших властивостей з теореми CAP. Деякі СУБД, наприклад Riak, дозволяють налаштовувати необхідні характеристики доступності-узгодженості навіть для окремих запитів шляхом завдання кількості вузлів, необхідних для підтвердження успіху транзакції.

Рішення NoSQL відрізняються як проектуванням з урахуванням масштабування. Іншими характерними рисами NoSQL-рішень є:

- Застосування різних типів сховищ.
- Можливість розробки бази даних без схеми.
- Лінійна масштабованість (додавання процесорів збільшує продуктивність).

Опис схеми даних у разі використання NoSQL-рішень може здійснюватись через використання різних структур даних: хеш-таблиць, дерев та інших.

Залежно від моделі даних та підходів до розподіленості та реплікації в NoSQL-руху виділяються чотири основні типи систем: «ключ – значення» (key-value store), «сімейство стовпців» (column-family store), документоорієнтовані (document store), графові.

Документоорієнтовані СУБД служать для зберігання ієрархічних структур даних і знаходять своє застосування у системах управління контентом та документальному пошуку. Приклади СУБД цього типу – CouchDB, Couchbase, Berkeley DB XML, MongoDB<sup>34</sup>.

Зокрема, MongoDB – документоорієнтована система управління базами даних, яка не вимагає опису схеми таблиць<sup>35</sup>. Вважається одним із класичних прикладів NoSQL-систем, використовує JSON-подібні документи та схему бази даних. Система підтримує ad-hoc-запити: вони можуть

---

<sup>34</sup> Seven Databases in Seven Weeks: A Guide to Modern Databases and the NoSQL Movement. Luc Perkins, Eric Redmond, Jim Wilson. – Pragmatic Bookshelf, 2018. – 360 pp.

<sup>35</sup> MongoDB: The Definitive Guide. Kristina Chodorow. – O'Reilly Media, 2013. – 432 pp.

повертати конкретні поля документів та користувацькі JavaScript-функції. Підтримується пошук за регулярними виразами. Можна також налаштувати запит на повернення випадкового набору результатів. Є підтримка індексів.

MongoDB підходить для наступних застосувань:

- реєстрація та зберігання інформації про події;
- системи управління документами, зокрема правовими;
- електронна комерція;
- дані моніторингу;
- мобільні додатки;
- сховище операційних даних та веб-сторінок.

### ***Питання для самоконтролю***

- Яка основна різниця між техніками та технологіями Big Data?
- Які сфери використання технік Big Data?
- Як можуть бути використані техніки та технології Big Data в сфері права?
- Які технології використовуються для збору та обробки даних з різних джерел в реальному часі?
- Які технології використовуються для зберігання великих обсягів даних?
- Які системи використовуються для розподіленої обробки даних?
- Чому консолідація даних є важливим інструментом для бізнесу та управління?
- Що таке пакетна аналітика?

- Які переваги має аналіз реального часу порівняно з пакетною аналітикою?
- Як забезпечити асинхронне отримання даних для гібридної аналітики?
- Як можна класифікувати дані за їх якістю?
- Що таке процеси ETL і яку функцію виконує кожен з етапів цих процесів?
- Які методи інтелектуального аналізу даних використовуються поряд із технікою класифікації?
- Як відрізнити класифікацію від кластеризації?
- Які кроки зазвичай включає процес класифікації?
- Які юридичні документи можна класифікувати за темою або категорією за допомогою моделей класифікації?
- Як машинне навчання може використовуватися для виявлення ризиків у правових питаннях?
- Які політичні дані можуть бути класифіковані за допомогою машинного навчання?
- Як класифікація ризиків корупції може допомогти виявити можливі порушення і прийняти відповідні заходи?
- Що таке кластеризація і як вона застосовується в машинному навчанні без вчителя?
- Які можуть бути переваги використання технік кластерного аналізу в праві та парламентському контролі?
- Які можуть бути результати кластерного аналізу нормативно-правових актів і як вони можуть бути корисні для парламенту?
- Що таке машинне навчання та які характеристики є його відмінностями від прямого розв'язання задач?

- Які типи машинного навчання розрізняють та в чому полягають їх відмінності?
- Поясніть задачу виявлення тематичних зв'язків між законопроектами?
- Які можуть бути переваги визначення тематичних зв'язків між законопроектами?
- Поясніть задачу виявлення незвичайної активності у соціальних мережах?
- Які можуть бути наслідки зловживанням правами на вираз своєї думки в соціальних мережах?
- Поясніть задачу виявлення аномалій у законодавчих актах?
- Які можуть бути наслідки неідентифікованих нетипових або підозрілих положень в законодавчих актах?
- Поясніть задачу зменшення розмірності законодавчих даних?
- Поясніть задачу класифікації депутатів у кластери за їх характеристиками?
- Що означає термін "NoSQL" і як він відрізняється від традиційних реляційних СУБД?
- Які проблеми вирішуються за допомогою NoSQL СУБД?
- Що означає термін "BASE" і яку важливу роль він грає в системах управління базами даних?
- Що таке теорема CAP і які властивості розподіленої системи можна гарантувати згідно з цією теоремою?

### ***3.3. Джерела Big Data***

Сьогодні головними джерелами Big Data є:

- Соціальні мережі, такі як Facebook, Twitter, Instagram, LinkedIn, YouTube та інші. Вони зберігають величезні

обсяги даних про користувачів, їх поведінку, зацікавлення, думки та багато іншого.

- Мобільні пристрої, такі як смартфони, планшети та носимі пристрої, що збирають дані про розташування, активність користувача та інші характеристики.
- Сенсорні пристрої, такі як датчики, GPS-приймачі, вимірювачі рівня шуму та інші, що збирають дані про навколишнє середовище, стан обладнання та інші характеристики.
- Електронні комунікації, такі як електронна пошта, месенджери, VoIP, що збирають дані про обмін повідомленнями та комунікації між користувачами.
- Великі бази даних, такі як реєстри, бази даних підприємств, організацій, установ, що збирають дані про їх діяльність та процеси.
- Медична та наукова інформація, що збирається зі статей, досліджень, публікацій та інших джерел.
- Інтернет та веб-джерела, що містять велику кількість інформації з різних галузей та сфер життя.

Загалом, джерела Big Data можуть бути дуже різноманітними, але вони мають одну спільну рису – вони збирають величезну кількість даних, які можуть бути використані для аналізу та виявлення закономірностей.

Узагальнюючи сказане, можна зробити висновок, що головні джерела Big Data сьогодні це:

- Інтернет речей (IoT) та підключені до нього пристрої;
- Веб-ресурси, соціальні мережі;

- дані компаній: транзакції, замовлення товарів та послуг, поїздки на таксі та каршерингу, профілі клієнтів.

Саме веб-ресурси та соціальні мережі розглядаються як ті, що мають відношення до парламентського контролю.

Розвиток напрямку Big Data історично пов'язаний із розвитком інформаційної складової мережі Інтернет, а саме, веб-ресурсів та соціальних мереж. Цим даним присутня багатоваріантність, різноманітність структур. Сьогодні прийнято виділяти три основні види даних за ступенем їхньої структурованості. Перший рівень – це звичні структуровані дані, які можуть бути представлені відокремленими та заздалегідь визначеними полями, наприклад, усі таблиці, які мають у певному полі заданої довжини заголовки, в іншому заздалегідь заданому полі – один із фактів, в іншому полі – інший із фактів, визначальних числові чи текстові значення семантичних змінних, які у заголовках. Структуровані дані добре зберігати в реляційних базах даних та керувати такими даними зручно, використовуючи спеціальну мову SQL – Structured Query Language. Ще недавно дуже поширені, такі дані визначають лише 10% всього обсягу згенерованих даних.

Другий рівень – це напівструктуровані чи слабоструктуровані (semistructured) дані. Дані такого типу не можуть бути представлені у вигляді таблиці через відсутність частини атрибутів у різних даних. Прикладом таких даних можуть бути файли у форматах SGML (Standard Generalized Markup Language), XML (eXtensible Markup Language) або BibTex, у яких немає певної схеми зберігання даних, але семантичний зміст різних елементів даних може бути визначений за аналізом вмісту самого файлу. Багато даних, що зберігаються в Web-просторі, відносяться до напівструктурованих, дані бібліографічних описів публікацій, наукові дані.



Нарешті, неструктуровані дані, які за визначенням не можуть підійти під описані вище види. Вони включають тексти, записані символами різних мов, записи звуків, зображення, відеофайли, повідомлення електронної пошти (текстова частина), твіти, презентації та інша бізнес-інформація поза вивантаженнями баз даних. Вважається, що від 80 до 90 відсотків усіх даних в організаціях належать до неструктурованих даних.

У відкритих джерелах та спеціалізованих базах даних, доступних у веб-просторі, міститься велика частина інформації, необхідної для проведення аналітичних досліджень, проте залишаються відкритими питання її знаходження та ефективного використання. При використанні веб-простору як найпотужнішого джерела Big Data, найважливішими є проблеми обсягу, навігації, наявності інформаційного шуму та динамічного характеру інформації в Інтернет.

Можливості доступу до інтернет-ресурсів, які приваблюють своєю відкритістю, обсягами та змістовною багатогранністю, на перший погляд здаються безмежними. Однак, незважаючи на такі якості, як відкритість та доступність, існуючу інфраструктуру веб-простору не можна визнати надійною та достовірною. Назвемо кілька проблем, властивих веб-простору:

- не вирішене завдання доступу користувачів до різноманітних веб-ресурсів з «одного вікна» для отримання узагальненого представлення потоків інформації з необхідної тематики;
- не забезпечено можливість своєчасного «нагадування» та «проштовхування» профільної для користувача інформації, що публікується на великій кількості веб-сайтів;

- досить велика ймовірність відмови в обслуговуванні критично важливих веб-ресурсів у найнеобхідніший час.

Сьогодні існують технології інтеграції контенту, які дозволяють частково вирішувати ці проблеми, забезпечуючи ефективний пошук і навігацію у веб-просторі, моніторинг та агрегацію відкритих веб-ресурсів. Для професійного пошуку та агрегації інформації з веб-простору використовуються технології Big Data, інформаційно-пошукові системи та сервіси.

В інформатиці на загальному рівні існують три фундаментальні поняття:

- дані;
- інформація;
- знання.

Поняття «знання» є найбільш складним і недостатньо вивченим. Досі для нього не існує загальноприйнятого визначення, і різні автори часто вкладають в нього різний зміст, залежно від загального контексту та своїх цілей. Але ми не матимемо потреби у його використанні, а тому не вдаватимось в існуючі теорії. Зазначимо лише, що знання передбачають фіксацію зв'язків між речами і явищами оточуючого нас світу, які дозволяють орієнтуватись в ньому і тією чи іншою мірою передбачати перебіг подій.

Зупинимось на понятті «дані». Звичайно його визначають як формалізоване представлення інформації у вигляді, яке придатне для передачі, зв'язку та обробки, і піддається багатократній інтерпретації. В такому розумінні воно до певної міри є вторинним по відношенню до поняття «інформація». Але інтенсивний розвиток інформаційних технологій, що спостерігається протягом останніх десятиріч,

змушує нас переглянути погляд на дані. А саме, на перший план виступає сприйняття даних як об'єкту застосування інформаційних технологій. В першу чергу мається на увазі процеси накопичення, збереження, обміну та машинної обробки. З цієї точки зору дані вже не обов'язково повинні бути пов'язані з конкретною інформацією, а також можуть взагалі не піддаватись жодній інтерпретації. На базовому рівні дані повинні мати лише одну властивість: відмінний від нуля обсяг.

На практиці використовуються два види організації даних:

- текстові дані;
- двійкові дані.

Текстові дані реалізуються як послідовності символів того чи іншого алфавіту (в узагальненому розумінні), а двійкові – як послідовності байтів.

В праві формальні послідовності символів як такі не розглядаються, оскільки з ними, взагалі кажучи, не пов'язані конкретні смислові структури. Тому для нас важливий певний тип організації даних, який може забезпечити наявність конкретного змісту в рамках заданої теми. А для цього дані перш за все повинні бути представлені як дискретні набори даних, кожний з яких асоціюється з тією чи іншою темою. Окрім іншого, це потрібно для того, щоб забезпечити зручний доступ до потрібного блоку даних.

В праві такі блоки містять не послідовності окремих символів, а синтаксичні конструкції, які мають конкретну інтерпретацію на визначеній мові. Як правило, такими конструкціями є фрази, але в ряді випадків можуть використовуватись окремі слова (наприклад, в тематичних словниках, онтологіях, тезаурусах тощо).

Важливо зрозуміти, що довільні послідовності не лише символів, але й слів чи фраз, взагалі кажучи, не можуть сприйматись як інформаційні блоки, щодо яких є сенс застосовувати інформаційні технології. Для цього потрібні певні набори даних, які мають визначені загальні характеристики, що фіксують їх конкретний зміст. Їх називають інформаційними одиницями. Тому не буде помилкою сказати, що інструментальні засоби інформаційних технологій в плані їх практичного застосування оперують інформаційними одиницями.

Існують дві основні категорії таких інформаційних одиниць:

- тексти;
- документи.

І нам важливо буде розрізнити їх, оскільки вони передбачають різні види машинної обробки, а отже і різні способи використання.

Текст є самодостатнім набором синтаксичних конструкцій, що не є чимось повністю цілісним і завершеним. Достатньо лише наявності смислової цілісності та внутрішньої структури. Повноцінний текст повинен мати початок і кінець але вони визначаються лише його змістом. Також характерною для тексту є його тема, яка може розпадатись на інші, вужчі теми (підтеми), і визначається його смисловою цілісністю.

Текст має визначений зміст, але сам по собі не передбачає якоесь конкретне використання. Він може містити в собі корисні відомості, але їх інтерпретація і пов'язані з нею практичні дії споживача не визначаються ні змістом, ні формою.

Але текст може бути доповнений додатковими характеристиками, внаслідок чого він набуває властивості,

що дозволяють його певним чином використовувати на практиці. Основні характеристики визначають спосіб, у який він оформлений. Часто додають те, що він повинен мати юридичну сили відповідно до чинного законодавства. Такий текст називають документом.

Документ – це текст, який має формально зафіксовані початок і кінець, а також набір атрибутів (заголовок, автор дата створення джерело тощо). Але головне те, що він від початку призначається для безпосереднього використання кінцевим споживачем з певною метою. Отже, в кінцевому рахунку можемо сказати, що документ – це текст, належним чином оформлений і створений з метою конкретного використання кінцевим споживачем. В цьому розумінні документ також може бути визначений як інформаційний продукт.

Документ має такі базові властивості:

- атрибутивність (наявність формалізованих складових, які не можуть бути з нього вилучені;
- функціональність (призначення для певного використання);
- структурованість (наявність системи зв'язків між його елементами, що надає йому завершену цілісність).

Виділяють такі ознаки документа:

- семантично окреслений зміст, тобто послідовно розміщені знаки, з яких складається документ, мають передавати його зміст;
- стабільний матеріальний носій, тобто фізична організація даних, що утворюють документ, повинна забезпечити можливість багаторазового використання та передачу інформації по лініях зв'язку;

- призначеність для використання в соціальній комунікації, тобто документ має бути спеціально створеним людиною для забезпечення певних комунікаційних цілей;
- завершення повідомлення, тобто документ повинен містити принаймні одне нефрагментоване повідомлення.

Документ виконує такі основні функції:

- інформаційна (зберігання та передача інформації);
- соціальна (задоволення соціальних потреб);
- культурна (забезпечення зв'язку між окремими особами та суспільними організаціями).

Важливою частиною документа є його реквізити. Вони, залежно від виду та призначення, можуть використовуватися по-різному, але наявність тих чи інших реквізитів є формальною ознакою документа.

Основними є такі реквізити:

- Державний герб;
- емблема організації;
- зображення нагород;
- код підприємства;
- код документа;
- назва міністерства чи відомства;
- назва підприємства;
- назва структурного підрозділу;
- індекс підприємства зв'язку, поштова адреса, номер телефону, факс у, номер рахунку в банку;
- назва виду документа;
- дата;
- індекс (вихідний номер документа);
- посилання на індекс і дату вхідного документа;
- місце складання чи видання;

- гриф обмеження доступу до документа;
- адресат;
- гриф затвердження;
- резолюція;
- заголовок до тексту;
- позначка про контроль;
- текст;
- позначка про наявність додатка;
- підпис;
- гриф погодження;
- візи;
- відбиток печатки;
- позначка про засвідчення копії;
- прізвище виконавця та номер його телефону;
- позначка про виконання документа й направлення його до справи;
- позначка про перенесення даних на машинний носій;
- позначка про надходження;
- запис про державну реєстрацію.

Текст документа має містити три основних складові:

- вступ;
- основна частина;
- закінчення.

В праві інформаційне ядро складають різноманітні документи, але існує також інформаційна периферія, що містить численні тексти, часто несистематизовані, розрізнені і фрагментарні. Тому повноцінна машинна обробка правової інформації повинна включати в себе роботу як з документами, так і з текстами.

***Питання для самоконтролю:***

- Які дані збирають соціальні мережі про користувачів?

- Які характеристики користувачів збираються за допомогою мобільних пристроїв?
- Які типи сенсорних пристроїв збирають дані про навколишнє середовище та стан обладнання?
- Які дані збираються за допомогою електронних комунікацій?
- Які типи баз даних зазвичай збирають дані про діяльність та процеси організацій?
- Які джерела медичної та наукової інформації зазвичай використовуються для збору даних?
- Які типи інформації можна знайти в інтернеті та веб-джерелах?
- Які є головні джерела Big Data зараз і чому саме вони вважаються такими важливими?
- Які види даних виділяють за ступенем їхньої структурованості?
- Які проблеми можуть виникнути при збереженні та обробці напівструктурованих даних?
- Які методи можна використовувати для зручного керування структурованими даними?
- Який відсоток згенерованих даних складають звичні структуровані дані?
- Які інформаційні проблеми виникають при проведенні аналітичних досліджень на веб-просторі?
- Які проблеми пов'язані з обсягом та навігацією веб-простору?
- Як можна вирішити проблему доступу користувачів до різноманітних веб-ресурсів з «одного вікна»?
- Чому не забезпечено можливість своєчасного «нагадування» та «проштовхування» профільної для користувача інформації?



- Які проблеми пов'язані з надійністю та достовірністю інфраструктури веб-простору?
- Які три фундаментальні поняття існують на загальному рівні в інформатиці?
- Чому поняття "знання" є найбільш складним і недостатньо вивченим з трьох понять?
- Які зв'язки між речами і явищами оточуючого світу фіксують знання?
- Яку роль відіграють знання в орієнтуванні в світі та передбаченні подій?
- Що таке документ в інформатиці?
- Які базові властивості має документ?
- Які ознаки документа виділяють фахівці?
- Які складові повинні бути формалізовані в документі?
- Які функції може виконувати документ в соціальній комунікації?

### **3.4. Бази даних**

Для здійснення парламентського контролю як одного із основних інструментів застосовуються правові бази даних. Правові бази даних – це електронні ресурси, що містять офіційні тексти законів, рішень судів, нормативно-правових актів та іншої правової інформації. Основна особливість правових баз даних полягає в тому, що вони містять структуровану та систематизовану інформацію з питань права, що дозволяє зручно шукати, отримувати та аналізувати її.

Окрім офіційних правових баз даних, існують також комерційні та некомерційні бази даних, що містять неофіційну правову інформацію, таку як коментарі, аналізи,

судові практики та інші дані. Вони також можуть бути корисними для використання у правовій інформатиці, але їх використання повинно бути обґрунтоване та враховувати можливі ризики, пов'язані з неофіційною інформацією.

Одна з особливостей правових баз даних полягає у вимогах до їх структури та формату, що дозволяє забезпечити їх зручне використання та обробку. Наприклад, для забезпечення структурованості та системності інформації в правових базах даних використовуються стандартизовані формати документів, такі як XML або JSON.

Іншою особливістю правових баз даних є необхідність оновлення та відповідності до законодавства, що забезпечує їх актуальність та надійність. Більшість офіційних правових баз даних оновлюються автоматично відповідно до змін у законодавстві, але в разі використання неофіційних баз даних, необхідно перевіряти їх актуальність та надійність.

Реферативні бази даних – це бази даних, що містять посилання на наукові, науково-популярні, публіцистичні матеріали та інші документи з різних джерел, такі як журнали, газети, збірники, монографії та інші документи.

Основною особливістю реферативних баз даних є наявність системи посилань на первинні джерела. Реферативні бази даних відображають не весь текст документів, а лише зведену інформацію про них, таку як назва, автор, дата, короткий опис, ключові слова та посилання на первинне джерело. Це дозволяє швидко знайти потрібну інформацію та ефективно працювати з великим обсягом документів.

Реферативні бази даних - це бази даних, що містять інформацію про певні видання (наприклад, наукові статті, книги, законодавство тощо) і забезпечують доступ до цієї інформації за допомогою різноманітних пошукових запитів.

Основна відмінність реферативних баз даних від правових баз даних полягає у тому, що реферативні бази даних зазвичай містять не сам текст документів, а його короткий опис або реферат.

Основні властивості реферативних баз даних:

- Стислість: інформація про документ зводиться до кількох речень або абзаців, що робить її більш зручною для огляду та вибіркового використання.
- Систематизованість: реферативні бази даних часто містять систематизовану інформацію про видання, що дозволяє користувачам знайти не тільки окремі документи, а й пов'язану з ними інформацію.
- Відкритий доступ: багато реферативних баз даних доступні для загального користування через Інтернет.
- Різноманітність форматів: реферативні бази даних можуть мати різноманітні формати, такі як електронні таблиці, бази даних на CD або DVD, онлайн-системи тощо.

Формати реферативних баз даних можуть бути різними, залежно від того, які джерела вони включають. Наприклад, реферативні бази даних можуть містити посилання на статті з наукових журналів, публікації з газет, реферати наукових досліджень, патенти, нормативні документи, судові рішення та інші види документів.

Однією з важливих властивостей реферативних баз даних є їх системність і структурованість. Реферативні бази даних мають чітко визначену структуру, що дозволяє швидко знаходити потрібну інформацію, використовуючи пошук за ключовими словами, авторами, датами та іншими параметрами. Крім того, реферативні бази даних постійно

оновлюються, що дозволяє отримувати актуальну інформацію.

Реферативні бази даних широко використовуються в правовій інформатиці, дозволяючи швидко знаходити потрібні нормативно-правові документи.

Деякі з відомих форматів реферативних баз даних: PubMed, Scopus, Web of Science, Google Scholar.

MARC (Machine-Readable Cataloging) – це один з найбільш відомих форматів реферативних баз даних, використовуваний для зберігання та передачі бібліографічних записів.

Машиночитаний каталог (MARC) – це стандарт, що використовується для опису бібліографічної інформації та організації записів у бібліотечних каталогах та бібліографічних базах даних. MARC був розроблений у 1960-х роках і став основою для розробки інших стандартів, таких як MARCXML і MARC21.

MARC складається з двох основних компонентів: формату запису та формату обміну. Формат запису використовується для створення записів у бібліографічній базі даних. Формат обміну використовується для передачі записів між бібліотеками та іншими організаціями, які займаються обробкою бібліографічної інформації.

Формат запису MARC складається з 3 частин: лідер, полів та підполів. Лідер містить загальну інформацію про запис, таку як тип запису, кількість полів та дата оновлення. Поля містять бібліографічну інформацію, наприклад автора, назву, видавництво тощо. Підполя використовуються для подробиць поля.

Однією з головних переваг формату MARC є те, що він дозволяє створювати структуровані записи з багатьма полями та підполями. Це робить його ідеальним для створення

бібліографічних баз даних з великою кількістю записів. Крім того, формат MARC дозволяє зберігати та обробляти більшість типів бібліографічної інформації, що робить його важливим інструментом для бібліотек та інших організацій, які працюють з бібліографічною інформацією.

Формат MARC продовжує застосовуватися в правовій сфері для створення та обміну бібліографічною та каталожною інформацією про правові документи. Він використовується для каталогізації законодавчих та інших правових документів у бібліотечних каталогах, в архівах та інформаційних системах.

Наприклад, у США більшість законодавчих бібліотек використовує формат MARC для каталогізації своїх колекцій. Також формат MARC використовується в низці спеціалізованих правових інформаційних систем, які надають доступ до баз даних з правової інформації, таких як Westlaw та LexisNexis.

Однак, з появою новіших технологій та форматів, таких як XML та JSON, застосування формату MARC у правовій сфері поступово скорочується. На заміну MARC вже приходять більш гнучкі та відкриті формати, які дозволяють ефективніше обробляти та обмінювати правову інформацію.

Наприклад, XML (Extensible Markup Language) - це формат, який дозволяє описувати структуру даних за допомогою тегів, і може використовуватись для створення різних типів баз даних. XML може бути використаний для передачі та обміну даними між різними програмами. Формат XML широко використовується в правовій сфері як стандарт для обміну даними між системами, програмами та документами. Використання XML дозволяє зберігати дані у структурованому форматі, що полегшує їх обробку та інтеграцію з іншими системами.

У правовій сфері формат XML використовується для створення та обміну різноманітної інформації, такої як законодавчі акти, судові рішення, звіти та інші документи. Також, XML може бути використаний для побудови різноманітних систем, які дозволяють автоматизувати процеси в правовій сфері, наприклад, систем управління документами, системи електронного документообігу, системи контролю доступу до інформації та інші.

Одним з прикладів використання формату XML в правовій сфері є стандарт Akoma Ntoso<sup>36</sup>, який використовується для представлення та обміну законодавчої інформації. Він був створений з метою покращення доступності, інтерпретації та обміну правовою інформацією.

Цей стандарт використовує XML для опису структури та змісту законодавчих документів, що дозволяє забезпечити їх автоматизовану обробку та інтеграцію з іншими системами. Akoma Ntoso ( Architecture for Knowledge Oriented Management of African Normative Texts using Open Standards and Ontologies ) — це міжнародний технічний стандарт для представлення документів виконавчої, законодавчої та судової влади в структурованому вигляді з використанням специфічного домену, юридичний словник XML. Термін akoma ntoso мовою акан у Західній Африці означає «зв'язані серця», і з цієї причини його було обрано для номінації цього правового стандарту XML. Звичайним аббревіатурою є AKN для позначення формату XML AKN .

Стандарт складається з:

---

<sup>36</sup> Флетт, Амелі; Лангнер, Арне; Leps, Olof (2022), "Phase I: Mapping Legal Concepts to Technical Objects", Model-Driven Development of Akoma Ntoso Application Profiles, Cham: Springer International Publishing, стр. 13–17, doi : 10.1007/978-3-031-14132-4\_3, ISBN 978-3-031-14131-7

- словника XML, який визначає відображення між структурою юридичних документів та їх еквівалентом у XML;
- специфікації схеми XML, яка визначає структуру юридичних документів у XML. Вони надають широкі можливості для опису кількох типів парламентських, законодавчих і судових документів, таких як законопроекти, акти та парламентські протоколи, судові рішення або вісники;
- рекомендованої угоди про найменування для надання унікального ідентифікатора правовим джерелам на основі моделі FRBR;
- визначення типу MIME.

Особливістю стандарту є його гнучкість і відкритість, яка дозволяє використовувати його для опису будь-якого виду правової інформації - від законодавчих актів до судових рішень. Крім того, АKN забезпечує зручний і однорідний доступ до правової інформації з різних джерел, що полегшує здійснення пошуку та аналізу правової інформації. Акوما Ntoso використовують у таких проектах, як Free Law Project, e-justice та багато інших. Стандарт є невід'ємною частиною руху відкритої правової інформації та відкритих даних у сфері права. Застосування стандарту Акوما Ntoso забезпечує зручну обробку та аналіз правової інформації за допомогою програмних засобів. Використання цього стандарту дозволяє підвищити ефективність обробки правової інформації та полегшити співпрацю між різними організаціями та установами у сфері права.

Інший формат, що став дуже популярним в останні роки - JSON (JavaScript Object Notation). Він також використовується для передачі даних між програмами, але відрізняється від XML тим, що використовує менше коду та є

більш читабельним для людей. JSON є відкритим форматом, що дає можливість багатьом програмам працювати з ним.

Окрім цього, існують інші формати, які можуть використовуватись для створення реферативних баз даних, такі як CSV (Comma-Separated Values), RDF (Resource Description Framework) та інші. Кожен з цих форматів має свої переваги та обмеження, і вибір формату залежить від конкретних потреб та вимог до зберігання та обробки даних.

***Питання для самоконтролю:***

- Що таке правові бази даних і яка їх основна функція?
- Які є особливості правових баз даних порівняно з іншими електронними ресурсами?
- Які види правових баз даних існують та яку правову інформацію вони містять?
- Які ризики пов'язані з використанням неофіційної правової інформації з комерційних та некомерційних баз даних?
- Які вимоги ставляться до структури та формату правових баз даних?
- Яка особливість правових баз даних, яка забезпечує їх актуальність та надійність?
- Як оновлюються офіційні правові бази даних?
- Яка роль посилань на первинні джерела в реферативних базах даних?
- Яка інформація міститься в реферативних базах даних?
- В чому полягає відмінність між реферативними та правовими базами даних?
- Яку важливу властивість мають реферативні бази даних?
- Які параметри дозволяють швидко знаходити потрібну інформацію в реферативних базах даних?



- Для яких галузей знань реферативні бази даних є широко використовуваними?
- Які формати реферативних баз даних є відомими, згадані в тексті?
- Що таке MARC і який він має статус у бібліотечній справі?
- З яких компонентів складається MARC і для чого вони використовуються?
- З якою метою був створений стандарт Akoma Ntoso?
- Яку особливість має стандарт Akoma Ntoso щодо опису правової інформації?
- В чому полягає відмінність JSON від XML?
- Які формати можуть використовуватись для створення реферативних баз даних?

### ***3.5. Метадані в правовій інформатиці***

Метадані – це дані про дані, тобто інформація, яка описує характеристики та властивості інших даних. В контексті правової інформатики метадані мають важливе значення, оскільки вони дозволяють ефективно збирати, організовувати, розшукувати та використовувати правову інформацію.

Метадані можуть містити інформацію про автора, дату створення, джерело, тему, тип та формат даних, правовий статус, рівень доступу, терміни, ключові слова та інші атрибути. Наприклад, в метаданих може бути зазначено, який законодавчий акт був ухвалений, коли це сталося, який орган ухвалив рішення, який був його зміст, а також яка версія закону чинна на даний момент.

Метадані в правовій інформатиці є важливим елементом для забезпечення якості та цілісності правової інформації. Вони допомагають зберігати, організовувати та управляти

правовою інформацією, забезпечуючи її легкий пошук та доступність користувачам. Метадані також використовуються для стандартизації форматів правової інформації, забезпечуючи сумісність та взаємодію між різними системами та джерелами правової інформації.

Бази даних метаданих є важливим інструментом для зберігання, керування та пошуку метаданих в правовій інформатиці. Метадані включають в себе інформацію про характеристики та властивості документа, такі як автор, дата створення, ключові слова, тема, номер статті тощо. Бази даних метаданих дозволяють організувати цю інформацію у структуровану форму, що полегшує її пошук та редагування.

У парламентському контролі бази даних метаданих можуть використовуватися для зберігання та організації документів, пов'язаних з різними аспектами діяльності парламенту, такими як законодавча діяльність, бюджетування, звітність тощо. Вони дозволяють швидко знаходити необхідні документи, забезпечують їх однорідність та зручний доступ до них. Крім того, бази даних метаданих можуть бути використані для автоматизованого аналізу та обробки даних, що сприяє підвищенню ефективності та точності парламентського контролю.

Однією з важливих особливостей баз даних метаданих є їхній структурований характер та можливість пошуку даних за різними критеріями. Це дозволяє швидко знаходити необхідні документи та інформацію та сприяє підвищенню ефективності роботи з даними. Крім того, бази даних метаданих зазвичай мають стандартизований формат та набір поля для введення даних, що забезпечує їхню однорідність та зручний доступ до них.

Для управління базами даних метаданих в правовій інформатиці можуть застосовуватись різні системи керування базами даних (СКБД). СКБД є програмними засобами для створення, організації та управління базами даних, які забезпечують ефективне зберігання, пошук, оновлення та видалення даних.

Однією з найпопулярніших архітектур СКБД є реляційна архітектура (SQL), яка реалізована в таких системах, як MySQL, Oracle та Microsoft SQL Server. Ця архітектура забезпечує зберігання даних у вигляді таблиць, які містять рядки та стовпці з конкретними значеннями. Такі СКБД забезпечують можливість проведення запитів до бази даних та забезпечують високий рівень безпеки.

Для більш спеціалізованого застосування можуть використовуватись інші типи СКБД, такі як NoSQL-бази даних, які забезпечують можливість зберігати дані в різних форматах, графові бази даних, які забезпечують зберігання та пошук даних за допомогою графів, та документ-орієнтовані бази даних, які забезпечують зберігання даних у вигляді документів

Застосування СКБД для метаданих в правовій інформатиці дозволяє зберігати, керувати та аналізувати метадані з різних джерел, забезпечує їх ефективний пошук та інтеграцію, а також дозволяє проводити аналіз та вивчення метаданих для здійснення парламентського контролю в правовій сфері.

#### ***Питання для самоконтролю:***

- Що таке метадані та яку інформацію вони містять?
- Яке значення метаданив мають в контексті правової інформатики?
- Які елементи метаданих важливі для забезпечення якості та цілісності правової інформації?

- Як метадані допомагають зберігати, організувати та управляти правовою інформацією?
- Як метадані використовуються для стандартизації форматів правової інформації?
- Яка основна перевага структурованого характеру баз даних метаданих?
- Які системи керування базами даних застосовуються в правовій інформатиці?
- Яка архітектура СКБД є найпопулярнішою в правовій інформатиці?
- Які можливості забезпечують СКБД з реляційною архітектурою?
- Який рівень безпеки забезпечують СКБД з реляційною архітектурою?

## **4. МЕРЕЖЕВІ МОДЕЛІ В ПРАВОВІЙ ІНФОРМАТИЦІ**

### ***4.1 Концепція складних мереж***

Концепція складних мереж або Complex Networks є теоретичним підходом до вивчення взаємозв'язків між елементами, які можуть мати складну структуру та багато рівнів взаємодії. Такі мережі можуть бути застосовані для моделювання різних систем, включаючи соціальні, економічні та технічні.

В правовій інформатиці концепція складних мереж може бути застосована для аналізу правових систем та їхніх зв'язків, зокрема для вивчення взаємозв'язків між законами, судовими рішеннями, прецедентами та іншими елементами правової системи. Крім того, такі мережі можуть допомогти відстежувати та аналізувати зміни в правовій системі з часом, зокрема зміни в інтерпретації законів та судової практиці.

Окрім того, концепція складних мереж може бути використана для аналізу взаємодії між різними агентами в правовій системі, такими як адвокати, судді, прокурори, правозахисники та інші учасники правового процесу. Аналіз таких взаємодій може допомогти зрозуміти динаміку правової системи та виявити потенційні проблемні місця, що потребують уваги та вдосконалення.

Таким чином, концепція складних мереж може бути корисним інструментом для аналізу та вдосконалення правової системи та взаємодії між її учасниками.

Основною причиною уваги до теорії складних мереж є їх важливість у багатьох сферах, таких як соціальні науки, біологія, економіка, фізика, інформаційні технології та ін. Складні мережі виникають у великій кількості природних і

штучних систем, де взаємодія між їх елементами може бути представлена у вигляді графів.

Теорія складних мереж дозволяє досліджувати та аналізувати взаємодію між елементами системи, виявляти ключові елементи та встановлювати їх вплив на систему в цілому. Це допомагає розуміти поведінку системи та прогнозувати її розвиток, що може бути корисним у різних галузях, включаючи правову інформатику.

В правовій інформатиці теорія складних мереж може застосовуватись, наприклад, для аналізу взаємодії між правовими нормами, їх ієрархії та впливу на розвиток правової системи в цілому. Крім того, складні мережі можуть бути використані для аналізу взаємодії між суб'єктами права та інших учасників правових відносин. В цілому, застосування теорії складних мереж може допомогти зрозуміти складність правових систем та покращити їх функціонування.

Концепція складних мереж може мати значення для парламентського контролю у багатьох аспектах. Основною роллю складних мереж у парламентському контролі є аналіз взаємозв'язків між різними джерелами даних та зборів інформації про ці зв'язки. Застосування аналізу складних мереж може допомогти виявити взаємозв'язки між різними аспектами законодавства, такими як закони, політики та програми дій, та оцінити їхній вплив на різні аспекти життя громадян.

Крім того, складні мережі можуть використовуватися для збору, обробки та аналізу даних про виборчі кампанії, результати виборів та різноманітні статистичні дані, що пов'язані з парламентським контролем. Це може допомогти парламентським комітетам та іншим органам здійснювати

більш об'єктивний та ефективний аналіз та контроль за виконанням законів та інших рішень.

Узагалі, застосування концепції складних мереж в правовій інформатиці може допомогти покращити розуміння та аналіз різноманітних законодавчих процесів та їх впливу на суспільство.

Разом з цим, практично усі сучасні мережі можна вважати складними. Так, наприклад, задача синтезу мереж допускає комбінаторний підхід, що спирається на представлення мережі у вигляді кінцевого графа, вершин якого відповідають вузлам мережі, а ребра – лініям зв'язку. В той же час, використання методів перерахування графів навіть для відносно невеликих мереж вважається неперспективним, оскільки необхідно дослідити величезну кількість можливих варіантів з'єднання вузлів лініями зв'язку.

Наприклад, у мережі з 10 вузлів існує  $2^{45}$  варіантів з'єднання. Для 10 вузлів теоретично можливо  $C_{10}^2 = \frac{10 \cdot 9}{2} = 45$  ліній сполучення. Кожна з цих можливих ліній зв'язку може реально існувати – стан «1», або не існувати – стан «0», тобто існує  $2^{45}$  можливостей.

Для меншої кількості вузлів (наприклад,  $n = 3$ ) варіанти можуть бути реально перебрані ( $2^{\frac{32}{2}} = 8$ ) (рис. 2).

### ***Параметри складних мереж***

Теорія складних мереж як область дискретної математики вивчає характеристики мереж, враховуючи не лише їх топологію, але й їх структуру, функції та поведінку, зокрема взаємодії між їх компонентами та різні процеси, які відбуваються на мережах. Застосування теорії складних мереж в правовій інформатиці може допомогти в аналізі

різноманітних законодавчих та регуляторних мереж, які виникають в процесі правотворення, та встановленні залежностей між ними. Наприклад, за допомогою аналізу складних мереж можна виявляти закономірності у процесах прийняття рішень та взаємодії між політичними акторами, які впливають на законодавчий процес.

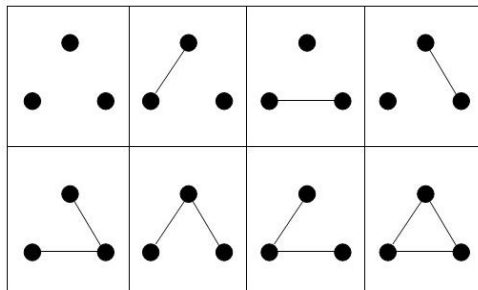


Рисунок 2 – Варіанти мереж при  $n = 3$

Найбільший внесок у розвиток теорії складних мереж внесли дослідники з різних галузей науки, включаючи фізику, математику, інформатику, соціологію та інші. Зокрема, до визначення основних характеристик складних мереж та їх аналізу внесли вагомий внесок дослідники Марк Гірціга, Альберт-Ласло Барабаши, Дункан Воттс, Стівен Строгац та інші. Вони досліджували різні аспекти складних мереж, такі як їх структуру, розмірність, робастність, моделювання та інші, що дозволило створити багато нових методів інтерпретації та аналізу складних мереж. Зокрема, дослідження соціальних мереж внесли значний внесок у розвиток теорії складних мереж. В соціальних мережах ми можемо виявити широкий спектр структур і поведінки, від простих до складних, і вони зазвичай мають велику кількість вузлів та зв'язків. Соціальні мережі також включають в себе різноманітність взаємодій між людьми, такі як дружба, співробітництво, конкуренція тощо. Це стало вихідним



матеріалом для вивчення складних мереж і розвитку нових методів аналізу та моделювання.

Дослідження соціальних мереж дозволили виявити різні структури мереж, такі як "малий світ", "безмежний масив", "багатомодульність" та інші. Ці структури можуть бути використані для розуміння складних систем в різних сферах, таких як біологія, економіка, технічні науки і, звичайно ж, правова інформатика. Вивчення соціальних мереж також дозволило розробити нові методики аналізу та класифікації мереж, що можуть бути застосовані для аналізу правових документів, дослідження законодавчої діяльності і парламентського контролю.

Термін «соціальна мережа» означає зосередження соціальних об'єктів, які можна розглядати як мережу (або граф), вузли якої – об'єкти, а зв'язки – соціальні стосунки. Цей термін було введено в 1954 році соціологом з «Манчестерської школи» Дж. Барнсом в роботі «Класи і збори в норвезькому острівному приході». У другій половині ХХ століття поняття «Соціальна мережа» стало популярним серед західних дослідників, при цьому як вузли соціальних мереж стали розглядати не лише представників соціуму, але і інші об'єкти, яким притаманні соціальні зв'язки. У теорії соціальних мереж отримав розвиток такий напрям, як аналіз соціальних мереж (Social Network Analysis, SNA). Сьогодні термін «соціальна мережа» має більш широке значення, ніж просто соціальний аспект, який він мав раніше. Соціальна мережа - це система взаємозв'язків між окремими сутностями, такими як особи, організації, групи тощо, що з'єднані між собою різними видами зв'язків, такими як дружба, співпраця, обмін інформацією, фінансові операції тощо. Соціальні мережі можуть бути представлені у вигляді графів, де вузлами є сутності, а зв'язки між ними - ребра графа.

Такі мережі можуть використовуватися в різних сферах, включаючи бізнес, медіа, науку та інше. Наприклад, в бізнесі соціальні мережі можуть використовуватися для аналізу взаємодії між клієнтами та компанією, що дозволяє розробляти більш ефективні стратегії маркетингу та збуту. В науці соціальні мережі використовуються для вивчення соціальних феноменів, таких як поведінка масової аудиторії в Інтернеті або розповсюдження епідемій.

У правовій інформатиці соціальні мережі можуть бути використані для вивчення структури та взаємодії різних суб'єктів права, таких як органи державної влади, правозахисні організації, адвокати, суди тощо. Аналіз соціальних мереж може допомогти виявити потенційні проблеми та нерівності у системі права, такі як нерівномірний доступ до правосуддя або корупційні схеми.

У рамках теорії складних мереж розглядають як статистичні, так і динамічні мережі, для розуміння структури яких необхідно враховувати принципи їх еволюції. Розглядання мереж як динамічних об'єктів може бути корисним для правової науки з декількох причин.

По-перше, в праві існують багато ситуацій, коли змінюються зв'язки між людьми, організаціями, інституціями і т.д. Наприклад, після виборів змінюється політична мережа, де кожен політик може мати різні зв'язки з іншими політиками. Динамічна модель може допомогти правовим дослідникам розуміти, які зміни відбуваються в мережі в часі, які впливи на ці зміни мають різні фактори і як ці зміни можуть вплинути на правову систему в цілому.

По-друге, аналіз динамічної мережі може допомогти виявити зміни відносин між різними агентами в мережі і зрозуміти, як ці зміни можуть вплинути на поведінку цих агентів. Це може бути особливо корисно в ситуаціях, коли

регуляторні органи спрямовують свої зусилля на забезпечення стабільності правової системи і мінімізацію негативних наслідків від змін в мережі.

По-третє, динамічні моделі мереж можуть бути використані для прогнозування змін в мережі, що може бути корисним для прийняття рішень у сфері права. Наприклад, правові дослідники можуть використовувати динамічну модель мережі для прогнозування того, які законодавчі зміни можуть виникнути в наступному році в залежності від змін в мережі агентів.

Отже, розглядання мереж як динамічних об'єктів може бути корисним для правової науки.

У теорії складних мереж виділяють декілька основних напрямів:

1. Статичні мережі - вивчення структури та властивостей мереж, які не змінюються в часі.
2. Динамічні мережі - вивчення мереж, які змінюються в часі, тобто змінюється структура взаємозв'язків між вузлами.
3. Мультирівневі мережі - вивчення мереж, які складаються з декількох рівнів, на кожному з яких взаємодіють окремі вузли.
4. Мережі з вагованими зв'язками - вивчення мереж, в яких зв'язки між вузлами мають різну вагу.
5. Мережі з різноманітними типами вузлів - вивчення мереж, в яких вузли можуть мати різні властивості та виконувати різні функції.
6. Мережі з кількома типами зв'язків - вивчення мереж, в яких можуть існувати різні типи зв'язків між вузлами.

У праві ці напрями можуть бути застосовані для аналізу структури та взаємодії різних юридичних осіб, таких як підприємства, установи, органи влади тощо, а також для вивчення динаміки змін у правовій системі та її впливу на соціальні процеси.

В правових дослідженнях можуть зустрічатись різноманітні структури складних мереж, залежно від об'єкту дослідження. Основні структури складних мереж, які можуть бути використані в праві, включають:

1. Графи відносин: це графи, що відображають взаємозв'язки між різними суб'єктами правових відносин, такі як зв'язки між підприємствами, громадянами, правовими органами тощо.
2. Графи аналізу тексту: це графи, що відображають взаємозв'язки між різними елементами тексту, такі як зв'язки між словами, реченнями або параграфами. Вони можуть використовуватися для аналізу правових документів, що дозволяє відстежити змістовні зв'язки між різними елементами цих документів.
3. Графи процесів: це графи, що відображають послідовність подій або процесів, що відбуваються у правових відносинах. Вони можуть бути використані для аналізу процесів вирішення правових питань, наприклад, процесу здійснення правосуддя.
4. Графи залежностей: це графи, що відображають залежності між різними елементами правової системи, такі як залежності між законами, судовими рішеннями або правовими актами. Вони можуть допомогти розуміти взаємозв'язки між різними правовими нормами та їхні взаємовпливи на правову систему в цілому.

Про «структуру співтовариства» в складній мережі можна говорити тоді, коли існує фрагмент мережі – група вузлів, які мають високу щільність ребер між собою, при тому, що щільність ребер між окремими фрагментами – низька. Традиційний метод для виявлення структури співтовариств – кластерний аналіз. Існують десятки прийнятних для цього методів, які базуються на різних мірах відстаней між вузлами, зважених шляхових індексах між вузлами тощо. Для великих соціальних мереж наявність структури співтовариств виявилася невід’ємною властивістю.

При аналізі складних мереж як і в теорії графів досліджуються параметри окремих вузлів; параметри мережі в цілому; мережеві підструктури.

### ***Параметри вузлів мережі***

У теорії складних мереж, параметрами вузлів можуть бути:

- Ступінь (degree) - кількість зв'язків, які виходять з вузла.
- Центральність (centrality) - міра впливу вузла на мережу.
- Близькість (closeness) - відстань від вузла до всіх інших вузлів в мережі.
- Ексцентричність (eccentricity) – найбільша з геодезичних відстаней (мінімальних відстань між вузлами) від вибраного вузла до інших;
- Посередництво (betweenness) – кількість найкоротших шляхів, що проходять через вибраний вузол;
- Коефіцієнт кластеризації (clustering coefficient) - міра того, наскільки вузли, що пов'язані з вузлом, також пов'язані між собою.

- Важливість (importance) - вагомість вузла в мережі залежно від його функціональної ролі або значення.
- Уразливість – рівень спаду ефективності (визначено нижче) мережі у разі вилучення вузла і усіх суміжних з ним ребер.
- Асортативність (assortativity) - тенденція вузлів в мережі зв'язуватися з вузлами зі схожими параметрами.

У правових дослідженнях можуть використовуватись різні параметри вузлів, залежно від конкретної задачі дослідження. Наприклад, для аналізу структури правових норм може використовуватись ступінь вузлів, а для виявлення ключових гравців у правовому процесі - центральність вузлів.

### *Загальні параметри мережі*

Деякі загальні параметри мережі, що можуть бути визначені в теорії складних мереж, включають:

1. Кількість вузлів: кількість вузлів або вершин у мережі.
2. Кількість зв'язків: кількість зв'язків або ребер у мережі.
3. Середня ступінь вузла: середнє значення кількості зв'язків, що має кожен вузол мережі.
4. Діаметр мережі: найбільша відстань між будь-якими двома вузлами у мережі.
5. Коефіцієнт кластеризації: міра того, наскільки сильно взаємодіють вузли у мережі.
6. Центральність вузла: міра того, наскільки важливий є кожен вузол у мережі.
7. Середня довжина шляху: середнє значення відстані між будь-якими двома вузлами у мережі.

8. Радіус мережі: мінімальна відстань між будь-яким вузлом і найвіддаленішим від нього вузлом у мережі.

Ці параметри можуть бути використані для опису структури мережі та її особливостей, що дозволяє дослідникам використовувати теорію складних мереж для вивчення різних явищ у праві та суспільстві.

Дослідження складних мереж є активною областю досліджень в сучасній науці. Деякі з актуальних завдань дослідження складних мереж на сьогоднішній день включають:

- Вивчення різноманітних моделей складних мереж. Створення нових теорій та алгоритмів для аналізу, моделювання та прогнозування поведінки мереж.
- Визначення фрагментів мережі (клік, кластерів), в яких вузли зв'язані між собою сильніше, ніж з вузлами з інших подібних фрагментів.
- Виділення фрагментів мережі (компонент зв'язності), які пов'язані усередині і не зв'язані або слабко зв'язані між собою.
- Знаходження перемичок, тобто вузлів, при вилученні яких мережа розпадається на незв'язані частини.
- Розробка нових методів та інструментів для аналізу складних мереж. Включає розробку нових алгоритмів виявлення спільнот, прогнозування розвитку мереж, класифікації вузлів мереж та інші.
- Дослідження взаємодії складних мереж з різноманітними системами. Наприклад, дослідження взаємодії соціальних мереж з економікою, правом та політикою.

- Дослідження різноманітних проблем, пов'язаних з мережами, включаючи проблеми керування, безпеки та приватності, а також проблеми пов'язані з використанням штучного інтелекту та машинного навчання.
- Застосування теорії складних мереж у різних областях, включаючи фізику, біологію, економіку, соціологію та інші.
- Розвиток нових технологій та методів, які дозволяють збирати дані про складні мережі з більшою точністю та швидкістю.

Ці завдання можуть бути застосовані в різних галузях, включаючи правову інформатику, де вони можуть допомогти в аналізі мереж взаємодії між людьми, розумінні динаміки правових систем та забезпеченні більш ефективного прийняття рішень.

### ***Розподіл степенів зв'язності вузлів***

Розподіл степенів вузлів у мережах - це статистичний показник, який вказує на те, скільки зв'язків (ребер) має кожен вузол мережі. Степінь вузла - це кількість ребер, які з'єднують його з іншими вузлами мережі. Функція розподілу степенів вузлів  $P(k)$ , яка визначається як ймовірність того, що вузол  $i$  має степінь  $k_i = k$ , тобто значення  $P(k)$  відповідає долі вершин із степенем  $k$ . Для орієнтованих мереж існує розподіл вхідних  $P^{in}(k^{in})$  і вихідних  $P^{out}(k^{out})$  напівстепенів, а також розподіл загального степеню  $P^{io}(k^{in}, k^{out})$ , що задає ймовірність знаходження вузла з вхідною  $k^{in}$  і вихідною  $k^{out}$  напівстепенями.

У мережах можуть існувати різні розподіли степенів вузлів. Мережі, що мають різні  $P(k)$ , демонструють дуже



різну поведінку.  $P(k)$  у деяких випадках може бути розподілами Пуассона ( $P(k) = e^{-m} m^k / k!$ , де  $m$  – математичне очікування), експоненціальним ( $P(k) = e^{-k/m}$ ) або степеневим ( $P(k) \sim 1/k^\gamma$ ,  $k \neq 0$ ,  $\gamma > 0$ ).

Наприклад, у деяких мережах кожен вузол має приблизно однакову кількість зв'язків, тобто розподіл степенів є рівномірним (uniform distribution). У інших мережах деякі вузли мають дуже багато зв'язків, тоді як інші мають дуже мало зв'язків. У таких мережах розподіл степенів вузлів є степеневим (power law distribution).

Розподіл степенів вузлів мережі визначає її структуру та властивості. Наприклад, у мережах з рівномірним розподілом степенів вузлів зв'язки розподілені рівномірно, що може вказувати на те, що ця мережа має однорідну структуру. Важливою особливістю багатьох реальних мереж є розподіл степенів вузлів  $P(k)$  за степеневим законом. Мережі із степеневим розподілом степенів зв'язності вузлів зветься безмасштабними (scale-free).

У мережах зі степеневим розподілом степенів вузлів кількість вузлів з високим ступенем (високий ступінь центральності) набагато менша, ніж вузлів з низьким ступенем. Це може свідчити про те, що в такій мережі існують головні вузли, які мають більше зв'язків з іншими вузлами і можуть відігравати важливу роль у функціонуванні мережі. При степеневому розподілі можливе існування вузлів з дуже високим ступенем, що, наприклад, практично не спостерігається в мережах з розподілом Пуассона.

У правовій інформатиці розподіл степенів вузлів може бути використаний для вивчення різних аспектів правової системи, таких як взаємодія між суддями, адвокатами, правозастосовними органами та іншими суб'єктами права.

Наприклад, можна вивчити розподіл ступенів вузлів у мережах правозастосовних органів, щоб з'ясувати, як часто вони співпрацюють між собою та як це впливає на результати їхньої діяльності. Крім того, розподіл ступенів вузлів може бути використаний для вивчення потенційних проблем в системі правосуддя, таких як надмірна концентрація влади в руках окремих суддів або недостатній рівень взаємодії між різними частинами правової системи.

Розподіл ступенів вузлів різними шляхами сформованих мереж також може бути використаний в задачі парламентського контролю для аналізу впливу окремих депутатів на прийняття рішення або для виявлення найбільш впливових фракцій чи комітетів. Наприклад, якщо в мережі, що відображає процес прийняття рішення, деякий депутат має високу ступінь вузла, то це може свідчити про його вплив на рішення. Аналогічно, якщо фракція має високу середню ступінь вузла, то це може свідчити про її вплив на рішення.

### ***Шлях між вузлами***

Шлях між вузлами є одним з основних параметрів мережі, який вказує на наявність зв'язку між двома вузлами. У теорії складних мереж це може бути використано для аналізу структури мережі, виявлення центральних вузлів, густини зв'язків, шляхів та клік. У правовій інформатиці шляхи між вузлами можуть бути використані для аналізу зв'язків між різними суб'єктами правових відносин, наприклад, між підприємствами, фізичними особами, установами і т. д. Використання параметра шляху між вузлами може допомогти виявити потенційно небезпечні зв'язки, конфлікти і відносини, що потребують уваги з боку правової системи. Також цей параметр може бути використаний для побудови систем підтримки прийняття рішень, які допомагають виявляти зв'язки між різними суб'єктами та прогнозувати можливі наслідки їхніх дій.

Якщо два вузли  $i$  та  $j$  можна з'єднати за допомогою послідовності з  $m$  ребер, то таку послідовність називають маршрутом між вузлами  $i$  та  $j$ , а  $m$  називають довжиною маршруту.

Говорять, що вузли  $i$  та  $j$  зв'язні, якщо існує маршрут між ними. Відношення зв'язності транзитивне, тобто якщо вузол  $i$  зв'язаний з вузлом  $j$ , а  $j$  зв'язний з  $k$ , то  $i$  зв'язаний з  $k$ .

При цьому маршрут, у якого початок і кінець знаходяться в одному і тому ж вузлі, а всі інші вершини використовуються лише один раз, називається циклом.

Відстань між вузлами визначається як довжина маршруту від одного вузла до іншого. Природно, вузли можуть бути сполучені прямо або опосередковано.

Шляхом між вузлами  $d_{ij}$  зветься найкоротша відстань між ними. Для всієї мережі можна ввести поняття середнього шляху, як середнє за усіма парами вузлів найкоротшої відстані між ними:

$$l = \frac{2}{n(n+1)} \sum_{i \geq j} d_{ij},$$

де  $n$  – кількість вузлів,  $d_{ij}$  – найкоротший шлях між вузлами  $i$  та  $j$ .

Угорськими математиками П. Ердешем (P. Erdős) і А. Реньї (A. Rényi) було показано, що середня відстань між двома вершинами у випадковому графові росте як логарифм від числа вершин.

Деякі мережі можуть виявитися незв'язними, тобто в них знайдуться вузли, відстань між якими є нескінченною. Відповідно, середній шлях може виявитися також

нескінченним. Для урахування таких випадків вводиться поняття глобальної ефективності мережі як середнього інверсного шляху між вузлами, що розраховується за формулою:

$$E = \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{d_{ij}},$$

де сума враховує усі пари вузлів. Ця характеристика відповідає ефективності мережі при пересилці інформації між вузлами (передбачається, що ефективність при пересилці інформації між двома вузлами  $i$  та  $j$  зворотно пропорційна відстані між ними).

Зворотна величина глобальної ефективності – середнє гармонійне геодезичних відстаней:

$$h = \frac{1}{E}.$$

Оскільки ця формула знімає проблему розбіжності при визначенні середнього шляху, то ця характеристика краще підходить для графів з декількома компонентами зв'язності. Ефективна відстань між двома вузлами в загальному випадку більша, ніж найкоротша відстань.

Мережі також характеризуються таким параметром як діаметр або максимальний по довжині шлях, тобто шлях, рівний максимальному значенню з усіх  $d_{ij}$ .

### ***Коефіцієнт кластеризації***

Д. Уаттс (D. Watts) і С. Стротац (S. Strogatz) визначили такий параметр мереж, як коефіцієнт кластеризації, який характеризує рівень зв'язності вузлів в мережі, тенденцію до утворення груп взаємозв'язаних вузлів, так званих клік (clique).

Коефіцієнт кластеризації є одним з основних параметрів мережі, що вказує на ступінь густоти взаємозв'язків між вузлами. Він показує, наскільки сильно взаємопов'язані вузли в мережі. Чим вищий коефіцієнт кластеризації, тим більш вірогідно, що вузли, що мають спільних сусідів, будуть також пов'язані між собою.

У праві кластерний аналіз може бути застосований для аналізу зв'язків між суддями, адвокатами, або іншими учасниками правової системи. Кластерний аналіз може допомогти виявити групи учасників, які мають більшість спільних зв'язків між собою, та встановити їх взаємодії. Також можна використовувати кластерний аналіз для аналізу зв'язків між суддями та рішеннями, які вони приймають, щоб виявити тенденції в їхній роботі та можливі упливи.

Крім того, кластерний аналіз може бути використаний для аналізу соціальних мереж, пов'язаних з правовою системою, наприклад, мережі адвокатів, які співпрацюють між собою в певних справах, або мережі громадських організацій, які працюють у сфері прав людини. Кластерний аналіз може допомогти виявити ключові вузли в таких мережах, а також встановити зв'язки між ними.

Для конкретного вузла коефіцієнт кластеризації показує, скільки найближчих сусідів цього вузла є також найближчими сусідами один одному. Коефіцієнт кластеризації для окремого вузла мережі визначається за наведеним нижче правилом. Нехай з вузла виходить  $k$  ребер, які сполучають його з  $k$  іншими вузлами, найближчими сусідами. Якщо припустити, що усі найближчі сусіди сполучені безпосередньо один з одним, то кількість ребер між ними складала б  $1/2 \cdot k(k-1)$ . Тобто це число, що відповідає максимально можливій кількості ребер, якими могли б з'єднуватися найближчі сусіди вибраного вузла.

Відношення реальної кількості ребер, які сполучають найближчих сусідів цього вузла до максимально можливого (такого, при якому усі найближчі сусіди цього вузла були б сполучені безпосередньо один з одним) називається коефіцієнтом кластеризації вузла  $i$  –  $C(i)$ . Природно, ця величина не перевищує одиниці.

Коефіцієнт кластеризації може визначатися як для кожного вузла, так і для усієї мережі. Відповідно, рівень кластеризації усієї мережі визначається як нормована по кількості вузлів сума відповідних коефіцієнтів окремих вузлів.

### ***Посередництво***

До втрати зв'язності в інформаційній мережі може привести розрив зв'язків між її компонентами, наприклад, при усуненні з інформаційного простору найбільш вагомих компонент (структурних компонент у разі мереж мови), тобто таких, які мають, найбільший коефіцієнт посередництва (betweenness). Цей коефіцієнт для конкретного вузла мережі визначається як сума по усіх парах вузлів мережі співвідношень кількості найкоротших шляхів між ними, що проходять через заданий вузол, до загальної кількості найкоротших шляхів між ними.

Значення вузла для мережі тим більше, чим в більшій кількості шляхів він задіяний. Тому, вважаючи, що обмін даними найчастіше відбувається по найкоротших шляхах, можна виміряти кількісно значення вузла з точки зору посередництва (betweenness), що визначається кількістю найкоротших шляхів що проходять через цей вузол. Ця характеристика визначає роль цього вузла у встановленні зв'язків в мережі – вузли з найбільшим посередництвом грають головну роль у встановленні зв'язків між іншими

вузлами в мережі. Посередництво  $b_m$  вузла  $m$  визначається за формулою:

$$b_m = \sum_{i \neq j} \frac{B(i, m, j)}{B(i, j)},$$

де  $B(i, j)$  – загальна кількість найкоротших шляхів між вузлами  $i$  та  $j$ ,  $B(i, m, j)$  – кількість найкоротших шляхів між вузлами  $i$  та  $j$ , що проходять скрізь вузол  $m$ .

Таким чином Betweenness - це міра того, наскільки часто вузол входить в найкоротший шлях між будь-якою парою вузлів у мережі. В правовій інформатиці цей параметр може бути застосований для визначення впливу конкретних вузлів в мережі на поширення інформації або передачу рішень. Наприклад, у судових мережах міра betweenness може вказувати на те, які судді мають значний вплив на передачу справ у вищі інстанції. Аналіз міри betweenness також може допомогти виявити затримки в передачі інформації або затримки в процесі прийняття рішень у правовій системі.

#### ***Питання для самоконтролю:***

- Що таке концепція складних мереж та які системи можуть бути модельовані за допомогою цієї концепції?
- Яким чином коефіцієнт кластеризації характеризує зв'язності вузлів в мережі?
- Які можливості надає кластерний аналіз для аналізу зв'язків між учасниками правової системи?
- Які соціальні мережі, пов'язані з правовою системою, можуть бути проаналізовані за допомогою кластерного аналізу?
- Яким чином кластерний аналіз може допомогти виявити ключові вузли в мережах?

- Які ще параметри мереж можуть бути використані для аналізу зв'язків між вузлами?
- Яким чином розраховується міра betweenness для вузлів в мережі, які можливі застосування міри betweenness в правовій інформатиці?
- Як може допомогти аналіз міри betweenness виявити затримки в передачі інформації в правовій системі?
- Чому міра betweenness є важливою в судових мережах?
- Які судді можуть мати значний вплив на передачу справ у вищі інстанції за допомогою міри betweenness?

#### ***4.2. Мережі мови в правовій інформатиці***

Першим кроком у застосуванні теорії складних мереж<sup>37,38</sup> до текстових документів є формування мережевої моделі цих документів у вигляді сукупності вузлів і зв'язків, тобто побудова мереж мови (Language Network)<sup>39</sup>, в яких виявляються найвагоміші вузли, іноді це так звані опірні слова або відповідні словосполучення.

У рамках теорії складних мереж розглядають як статистичні, так і динамічні мережі, для розуміння структури яких необхідно враховувати принципи їх еволюції. Розглядання мереж як динамічних об'єктів може бути корисним для правової науки з декількох причин.

По-перше, в праві існують багато ситуацій, коли змінюються зв'язки між людьми, організаціями, інституціями і т.д. Наприклад, після виборів змінюється політична мережа, де кожен політик може мати різні зв'язки з іншими

---

<sup>37</sup> Strogatz S.H. Exploring Complex Networks // Nature, 2001. – 410. – P. 268-276.

<sup>38</sup> Albert R., Barabasi A.-L. Statistical mechanics of complex networks // Reviews of Modern Physics, 2002. – 74. – P. 47.

<sup>39</sup> Головач Ю., Пальчиков В. Лис Микита і мережі мови, Журн. Фіз. Досл., 2006. – № 10. – С. 247-291.



політиками. Динамічна модель може допомогти правовим дослідникам розуміти, які зміни відбуваються в мережі в часі, які впливи на ці зміни мають різні фактори і як ці зміни можуть вплинути на правову систему в цілому.

По-друге, аналіз динамічної мережі може допомогти виявити зміни відносин між різними агентами в мережі і зрозуміти, як ці зміни можуть вплинути на поведінку цих агентів. Це може бути особливо корисно в ситуаціях, коли регуляторні органи спрямовують свої зусилля на забезпечення стабільності правової системи і мінімізацію негативних наслідків від змін в мережі.

По-третє, динамічні моделі мереж можуть бути використані для прогнозування змін в мережі, що може бути корисним для прийняття рішень у сфері права. Наприклад, правові дослідники можуть використовувати динамічну модель мережі для прогнозування того, які законодавчі зміни можуть виникнути в наступному році в залежності від змін в мережі агентів.

Поряд з послідовним аналізом текстових документів, побудова мереж, вузлами яких є такі елементи, як слова або словосполучення, тобто фрагменти природної мови, дозволяє виявляти структурні елементи текстів, без яких тексти втрачають свою зв'язність.

Мережі мови є важливим інструментом в правовій інформатиці, оскільки дозволяють аналізувати структуру та взаємозв'язки між словами, термінами та поняттями у правових документах та текстах.

Один з методів аналізу мереж мови - це побудова графів, де вершинами є слова, а ребрами - зв'язки між ними (наприклад, згадування одного слова в тому ж реченні з іншим словом). Такий граф може бути використаний для виявлення ключових термінів, що використовуються в

конкретній галузі права, а також для аналізу семантичних відносин між словами та поняттями.

Також мережі мови можуть використовуватись для аналізу правових документів та розуміння їх структури. Наприклад, можна використовувати алгоритми класифікації текстів для автоматичного розподілу документів за темами або типами документів.

Інший метод аналізу мереж мови - це використання моделей векторного простору, які дозволяють відобразити слова у векторному просторі та вимірювати семантичну відстань між ними. Це може бути використано для автоматичної кластеризації слів за їх семантикою, а також для пошуку синонімів та антонімів у правових текстах.

Одним з важливих напрямків дослідження мереж мови є аналіз судових рішень та їх використання для виявлення судової практики та розуміння правових термінів та понять.

Для аналізу судових рішень та розуміння правових термінів та понять у мережах мови в правовій інформатиці можуть бути використані наступні методи:

- Аналіз частоти вживання слів та їх зв'язків у текстах судових рішень.
- Побудова графів залежностей між словами в текстах судових рішень для виявлення семантичних зв'язків та класифікації рішень за подібністю.
- Використання алгоритмів машинного навчання для класифікації судових рішень та автоматичної обробки правової інформації.
- Аналіз спільнот слів та тематичного моделювання для виявлення ключових тем та понять, що використовуються у судових рішеннях.

- Побудова мережі мови, яка відображає зв'язки між словами та їх вживання в правових текстах. Ця мережа може бути використана для аналізу термінології та структури правових текстів.

Відомо декілька підходів до побудови мереж з текстів і різні способи інтерпретації вузлів і зв'язків, що приводить, відповідно, до різних видів представлення таких мереж. Вузли можуть бути сполучені між собою, якщо відповідні їм слова стоять поряд у тексті<sup>40,41</sup>, належать до одного речення або абзацу<sup>42</sup>, сполучені синтаксично<sup>43</sup> або семантично<sup>44</sup>.

Збереження синтаксичних зв'язків між словами призводить до представлення тексту у вигляді спрямованої мережі (Directed Network), де напрям зв'язку відповідає підпорядкуванню слова.

Якщо поставити у відповідність кожному слову вузол мережі і з'єднати кожні два вузли зв'язком тоді, коли відповідні ним слова стоять у реченні поруч, то таке представлення називають *L*-простором. У *L*-просторі, так само як і в інших наведених нижче мережевих моделях, при виникненні кратних зв'язків прийнято зберігати лише один з них.

---

<sup>40</sup> Ferrer-i-Cancho, R., Sole R.V. The small world of human language // Proc. R. Soc. Lond. B 268, 2261 (2001).

<sup>41</sup> Dorogovtsev S.N., Mendes J. F. F. Language as an evolving word web // Proc. R. Soc. Lond. B 268, 2603 (2001).

<sup>42</sup> Caldeira S.M.G., Petit Lobao T.C., Andrade R.F.S., Neme A., Miranda J.G.V. The network of concepts in written texts // Preprint physics/0508066 (2005).

<sup>43</sup> Ferrer-i-Cancho, R. The variation of Zipf's law in human language. // Phys. Rev. E 70, 056135 (2005).

<sup>44</sup> Motter A.E., de Moura A.P.S., Lai Y.-C., Dasgupta P. Topology of the conceptual network of language // Phys. Rev. E 65, 2002. –065102(R).

Традиційно розрізняють чотири різновиди мереж мови (просторів):

1. *L*-простір. Зв'язуються сусідні слова, які належать до одного речення. Кількість сусідів для кожного слова (вікно слова) визначається радіусом взаємодії  $R$ , найчастіше розглядається випадок  $R = 1$ .

2. *B*-простір. Розглядаються вузли двох видів, відповідні реченням і словам, що належать до них.

3. *P*-простір. Усі слова, які належать до одного речення, зв'язуються між собою.

4. *S*-простір. Речення зв'язуються між собою, якщо у них застосовуються однакові слова.

У випадку *L*-простору зв'язки можуть враховувати не лише «найближчих сусідів», але і групи з декількох слів, які знаходяться на певній відстані один від одного. Для цього вводиться поняття «радіусу дії»  $R$ : при  $R = 1$  зв'язок існує лише між найближчими сусідами, при  $R = 2$  – між найближчими і наступними близькими сусідами і т. д. Змінна  $R$  може приймати значення від  $R = 1$  до  $R_{\max}$ , де  $R_{\max} + 1$  – загальна кількість слів у реченні. Зростання «радіусу взаємодії»  $R$  у цьому випадку призводить до зростання кількості зв'язків, досягаючи насичення при  $R = R_{\max}$ .

Ще один спосіб представити текст у вигляді мережі полягає у використанні дводольних (bipartite) графів. У такому представленні (*B*-простір) розглядаються вузли двох видів. Один вид відповідає реченням, другий – словам. Зв'язок між різними вузлами означає, що слово належить реченню.

У *P*-просторі усі слова, що належать одному реченню, вважаються зв'язаними між собою.

У  $S$ -просторі вузли відповідають реченню, а зв'язок між вузлами-реченнями встановлюється у тому випадку, якщо у відповідних реченнях є загальні слова.

Для мережі, побудованої на підставі Британського національного корпусу ( $L$ -простір мови,  $R = 1$ ) виявилось, що ця мережа англійської мови безмасштабна, а поведінка степеню  $P(k)$  характеризується двома режимами степеневого розподілу із значеннями відповідних степеневих показників  $\gamma = 1,5$  для  $k < 2000$  і  $\gamma = 2,7$  для  $k > 2000$ .

Згідно з визначенням, якщо середня довжина найкоротшого шляху росте з розміром (кількістю вузлів) мережі повільніше за будь-яку степеневу функцію, то мережа є «малим світом». Мережі малого світу надзвичайно компактні. Для згаданої вище мережі англійської мови довжина найкоротшого шляху складає лише  $\langle l \rangle = 2,63$ . Оскільки зростання  $R$  призводить лише до додавання нових зв'язків, то  $\langle l \rangle$  зменшуються із зростанням  $R$ .

Специфічною формою кореляції в мережах є утворення кластерів. Коефіцієнт кластеризації  $C$  характеризує схильність мережі до утворення сполучених трійок вузлів. Відомо, що для повного графа  $C = 1$ , а для мережі у формі дерева  $C = 0$ .

Відношення середнього коефіцієнта кластеризації досліджуваних мереж до коефіцієнта кластеризації класичного випадкового графа свідчить про те, що мережі мови є добре корельованими структурами. Такі кореляції ростуть із зростанням «радіусу взаємодії»  $R$ .

Для Британського національного корпусу на підставі аналізу текстів, які містили  $\approx 10^7$  слів, набуто значення коефіцієнта кластеризації  $\langle C \rangle = 0,687$ .

У випадку розгляду  $P$ -простору кожне слово-вузол пов'язано з усіма іншими словами, які належать спільному реченню. Таким чином, кожне речення тексту входить в мережу як повний граф – кліка взаємозв'язаних вузлів. Різні речення-кліки об'єднуються в мережу завдяки загальним словам. У  $L$ -просторі слова зв'язуються в межах вікна, розмір якого характеризується величиною  $R$ . Коли розмір вікна  $R$  стає рівним розміру речення, то представлення цього речення в  $L$ - і в  $P$ -просторах співпадають. Відповідно, коли розмір вікна стає рівним розміру найбільшого речення тексту ( $R = R_{\max}$ ), то представлення усього тексту в  $L$ - і в  $P$ -просторах співпадають.

На практиці доведено, що мережа мови є сильно корельованим безмасштабним малим світом (Scale-Free Small World). Існує ряд праць, в яких зроблена спроба пояснити властивості мереж мови за допомогою сценарію переважного приєднання (Preferential Attachment<sup>45</sup>), розглядаючи їх як результат процесу зростання, коли нові вузли-слова з більшою ймовірністю приєднуються до вузлів-хабів, що мають багато зв'язків.

Мережі мови можуть бути корисним інструментом для парламентського контролю при вирішенні низки завдань, серед яких:

1. Аналіз законопроектів: Мережі мови можуть допомогти аналізувати законопроекти, які розглядаються парламентом. Застосовуючи методи аналізу тексту, можна виділити ключові терміни, теми та поняття в законопроектах та досліджувати зв'язки між ними. Це може допомогти парламентаріям зрозуміти зміст

---

<sup>45</sup> Albert R., Jeong H., Barabasi A.-L. Diameter of the world wide web // Nature (London), 1999. – 401, 130.

законопроектів та їх взаємозв'язки з іншими законами та політичними процесами.

2. Моніторинг медіа: Мережі мови можуть бути використані для моніторингу медіа, щоб виявляти ключові теми та поняття, які стосуються роботи парламенту. Аналізуючи статті, новини та інші джерела інформації, можна отримати уявлення про те, яким чином громадськість сприймає роботу парламенту та які проблеми є для неї важливими.
3. Аналіз дебатів: Мережі мови можуть допомогти аналізувати дебати, що відбуваються в парламенті. Застосовуючи методи аналізу тексту та мереж мови, можна визначити, які питання та теми є найбільш важливими для депутатів, як вони пов'язані між собою, та які депутати активніше беруть участь у дискусіях.
4. Моніторинг соціальних мереж: Мережі мови також можуть бути використані для моніторингу соціальних мереж, щоб виявляти громадську думку про певні законопроекти, питання або депутатів.
5. Аналіз структури законодавчих актів: Мережі мови можуть допомогти виявити структуру законодавчих актів, візуалізувати їх зв'язки та ієрархію та допомогти у зрозумінні змісту та цілей законодавства.
6. Пошук законодавчої інформації: Мережі мови можуть бути використані для розробки систем пошуку інформації, що дозволяють швидко знайти закони, накази та інші законодавчі документи, пов'язані з певною темою або проблемою.
7. Виявлення важливих тем та понять: Мережі мови можуть допомогти виявити важливі теми та поняття в законодавстві, що можуть бути важливими для

парламентського контролю. Такі теми та поняття можуть бути візуалізовані у вигляді графіків та діаграм.

8. Аналіз зв'язків між законодавчими актами: Мережі мови можуть бути використані для виявлення зв'язків між законодавчими актами, що дозволяє зрозуміти взаємодію між різними законами та їх вплив на практику застосування законодавства.
9. Моніторинг впливу законодавства: Мережі мови можуть допомогти у моніторингу впливу законодавства на різні галузі та аспекти життя громадян. Вони можуть допомогти виявити позитивні та негативні наслідки законів та допомогти у покращенні якості законодавства.

Отже, мережі мови можуть допомогти в підвищенні ефективності парламентського контролю та покращенні якості законодавства.

***Питання для самоконтролю:***

- Які переваги має застосування теорії складних мереж до аналізу текстових документів?
- Що таке мережева модель документів і як вона формується?
- Які принципи еволюції динамічних мереж необхідно враховувати для розуміння їх структури?
- Які ситуації в праві можуть бути розглянуті за допомогою динамічних моделей мереж?
- Які можуть бути користі від аналізу змін відносин між агентами в мережі з метою стабілізації правової системи?
- Які можливості відкриває аналіз динамічної моделі мереж для прогнозування змін в мережі агентів та прийняття рішень у сфері права?



- Яким чином можна використати граф, побудований за допомогою методу аналізу мереж мови, для виявлення ключових термінів в галузі права?
- Які алгоритми класифікації текстів можуть бути використані для автоматичного розподілу правових документів за темами або типами документів?
- Яким чином моделі векторного простору можуть допомогти в автоматичній кластеризації слів за їх семантикою у правових текстах?
- Яким чином можна використовувати аналіз судових рішень для виявлення судової практики та розуміння правових термінів та понять?
- Які різновиди мереж мови традиційно розрізняють?
- Які особливості L-простору мережі мови?
- Як визначається радіус взаємодії для L-простору мережі мови?
- Які вузли розглядаються в V-просторі мережі мови?
- Які зв'язки між вузлами в V-просторі мережі мови означають, що слово належить реченню?
- Які методи аналізу тексту використовуються для виділення ключових термінів, тем та понять у законопроектах?
- Які джерела інформації можуть бути використані для моніторингу медіа, щоб виявляти ключові теми та поняття, які стосуються роботи парламенту?
- Як можуть бути використані мережі мови для моніторингу соціальних мереж та виявлення громадської думки про законопроекти, питання або депутатів?
- Які методи візуалізації можуть бути використані для подання важливих тем та понять в законодавстві у вигляді графіків та діаграм?

### ***4.3. Корпусна лінгвістика в правовій інформатиці***

Корпусна лінгвістика – це той розділ мовознавства, що вивчає створення, обробку та використання текстових корпусів. В рамках комп'ютерної лексикографії словники часто формуються шляхом сканування і подальшої обробки паперових словників, проте на цей час все частіше початковим матеріалом для отримання необхідної лексики є текстові корпуси.

Корпусна лінгвістика - це підхід до вивчення мови, який ґрунтується на аналізі великих обсягів текстових даних, зібраних з різних джерел і додатково анотованих та інтерпретованих для отримання різноманітних лінгвістичних даних.

У правовій інформатиці корпусна лінгвістика може застосовуватися для аналізу правових документів, зокрема законодавчих текстів, судових рішень, договорів тощо. Завдяки такому аналізу можна виявляти закономірності в використанні певних термінів, понять та конструкцій у правових текстах, виявляти недоліки та неузгодженості у законодавстві та розробляти рекомендації щодо його удосконалення.

Крім того, корпусна лінгвістика може використовуватися для автоматичної обробки та аналізу текстів, що сприяє автоматизації процесів збору та аналізу даних, зменшенню ризику помилок та збільшенню ефективності роботи правових органів та юридичних фірм.

Текстовим корпусом у лінгвістиці називають сукупність текстів, зібраних відповідно до певних принципів, розмічених за відповідними правилами (морфологічною, акцентною, синтаксичною і іншою розміткою) і, як правило, доповнених пошуковою системою. Іноді корпусом ("корпус першого порядку") називають просто будь-які колекції текстів,

об'єднаних якоюсь загальною ознакою (мовою, жанром, автором тощо). Найпростіший лінгвістичний корпус (корпус першої позиції) як будь-яка колекція текстів з певної тематики, які є доступними в електронній формі. Виходячи з цього визначення, база даних «Законодавство України» (<http://zakon.rada.gov.ua>), що містить нормативно-правові документи України вже є лінгвістичним текстовим корпусом, нехай і найпростішим за структурою.

Разом з цим, уявленням автора відповідає визначення: «Під лінгвістичним корпусом текстів розуміється великий, уніфікований, структурований, розмічений, філологічно компетентний масив мовних даних, представлений в електронному вигляді й призначений для вирішення різних лінгвістичних завдань»<sup>46</sup>.

Нині існують сотні різних текстових корпусів для різних мов, за різною тематикою, з різним рівнем розмітки. Розмічені текстові корпуси створюються і використовуються як для лінгвістичних досліджень, так і для налаштування (навчання) моделей і процесорів за допомогою відомих математичних методів машинного навчання. Так, машинне навчання застосовується для налаштування методів вирішення лексичної неоднозначності, анафоричних посилань, розпізнавання частини мови<sup>47</sup>.

Останнім часом все частіше як найповніший лінгвістичний ресурс розглядаються тексти з мережі Інтернет, який стає найрепрезентативнішим джерелом зразків сучасної мови, проте його використання як корпусу вимагає деяких, обмежень, розробки спеціальних технологічних засобів.

---

<sup>46</sup> Ланде Д.В. Елементи комп'ютерної лінгвістики в правовій інформатиці. – К.: НДІП НАПрН України, 2014. – 154 с.

<sup>47</sup> Широков В.А., Бугаков О.В., Грязнухіна Т.О. Корпусна Лінгвістика. – К.: Довіра, 2005. – 471 с.

Доцільність створення текстових корпусів пояснюється:

- представленням лінгвістичних даних в реальному контексті;
- великою повнотою даних (при великому обсязі корпусу);
- можливістю багатократного використання корпусу, створеного один раз, для вирішення різних лінгвістичних завдань.

Робота з текстовими корпусами на цей час є одним з основних методів лінгвістичних досліджень. Ще у 1960-і роки було створено Брауновський корпус (США), який включає 1 млн. слів; у 1970-і було створено корпус LOB (Великобританія, Норвегія), який також включає 1 млн. слів. У 1980-і роки почали створюватися такі корпуси, як: Машинний Фонд російської мови, Уппсальський корпус російської мови (Швеція), корпус The Bank of English, Birmingham. У 1990-і створено British National Corpus, який включає 100 млн. слів, а також інші національні корпуси (угорський, італійський, хорватський, чеський, японський). На цей час корпус The Bank of English, Birmingham включає 600 млн. слів. На початку XXI ст. створювалися такі корпуси, як American National Corpus, 100 млн. слів і Gigaword corpora (англійський, арабський, китайський), що включає 1 млрд. слів.

Основними параметрами лінгвістичного текстового корпусу є:

- Розмір – кількість текстів, що містяться в корпусі.
- Характер текстів – типи текстів, що містяться в корпусі, такі як юридичні документи, наукові статті, публіцистичні тексти та інші.

- Джерела – джерела текстів, з яких складається корпус, такі як газети, журнали, веб-сайти, юридичні бази даних і т.д.
- Мови – мови, на яких написані текстові документи в корпусі.
- Тематика - тематична область текстів в корпусі, наприклад, право, медицина, біологія та інші.
- Авторство – автори текстів в корпусі, якщо це відомо.
- Дата створення – дата створення текстів в корпусі.
- Рівень обробки – рівень лінгвістичної обробки текстів в корпусі, наприклад, рівень токенизації, стемінгу, лематизації тощо.
- Формат – формат зберігання текстових документів в корпусі, наприклад, текстові файли, бази даних, XML-файли та інші.
- Доступність – доступність корпусу для використання, наприклад, відкритий доступ, доступ лише після реєстрації або доступ лише за плату.

Виходячи з цього, лінгвістичний текстовий корпус повинен бути достатньо великого обсягу; бути структурованим або розміченим; бути наведеним в електронному вигляді; доповнюватись спеціальним програмним забезпеченням для роботи з ним.

Цінність текстового лінгвістичного корпусу вбачається в наступному:

- одного разу зроблений корпус може використовуватися при рішенні різних задач;
- корпус містить мовні дані в їх реальному оточенні, що дозволяє досліджувати лексичну і граматичну структуру

мови, а також безперервні процеси мовних змін, що відбуваються впродовж певного часу;

- корпус характеризується показністю, або збалансованим складом текстів, що дозволяє використовувати його для тестування інформаційно-пошукових систем, відповідних морфологічних модулів, систем перекладу тощо;
- корпус може використовуватися при вивченні мови, оскільки за допомогою корпусу можна перевіряти особливості вживання мовних одиниць;
- корпусом можуть користуватися перекладачі. За наявності сумнівів про те, яким чином слід перекласти ту або іншу фразу, можна задати запит пошуковій системі корпусу і порівняти частоту вживань наявних варіантів.

Повнотекстовий пошук у типовому лінгвістичному текстовому корпусі на цей час доповнено такими параметрами, як:

- урахування порядку слів;
- пошук у вибраній підмножині об'єктів;
- використання нормалізації слів;
- підключення синонімічних ланцюжків до слів;
- урахування морфологічних і граматичних параметрів слів.

Розмітка (приписування тексту певної інформації для зручнішого аналізу) – суттєва характеристика сучасного текстового корпусу, яка відрізняє його від електронних колекцій і енциклопедій.

Існують різні типи розмітки:

- метатекстова (автор, назва, обсяг, тематика тощо), що характеризує текст у цілому;
- структурна розмітка є інформацією щодо структури тексту, яка дозволяє відокремити одне слово від іншого, виділити межі словосполучення, речення, тексту;
- лінгвістична розмітка, яка полягає в приписуванні одиницям тексту певної лінгвістичної інформації.

Вважається, чим багатша і різноманітніша розмітка, тим вищою є наукова і навчальна цінність корпусу.

Простір електронних текстових корпусів дає можливість результативного створення і використання електронних конкордансів.

Електронний конкорданс - це комп'ютерна програма, що дозволяє швидко і ефективно знаходити всі випадки вживання певного слова або фрази в текстовому корпусі. У правовій інформатиці електронні конкорданси використовуються для знаходження відповідних законодавчих норм, судових рішень, коментарів до законів тощо.

Основні переваги використання електронних конкордансів включають:

1. Швидкий та точний пошук: з використанням електронного конкордансу можна швидко знайти всі випадки вживання певного слова або фрази в текстовому корпусі, що значно економить час та збільшує точність пошуку.
2. Аналіз структури тексту: електронний конкорданс дозволяє аналізувати структуру тексту, зокрема, вживання певних слів та їх контекст.

3. Побудова колокацій: з використанням електронного конкордансу можна побудувати колокації - словосполучення, що часто вживаються разом у певному контексті.
4. Порівняння законодавчих норм: електронний конкорданс дозволяє порівняти різні версії законодавчих норм та знайти різниці в їх формулюванні.
5. Аналіз судової практики: з використанням електронного конкордансу можна аналізувати судову практику та знаходити приклади вирішення певних правових питань.
6. Розробка програмного забезпечення: на основі даних, отриманих з електронного конкордансу, можна розробляти програмне забезпечення для автоматизації правової діяльності.

Конкорданси відкривають перспективи моделювання мовної картини світу на основі статистичної властивості мови, що проявляються лише на великих обсягах інформації та знаходять своє відображення в лінгвістичних текстових корпусах.

Таким чином, використання електронних конкордансів в правовій інформатиці дозволяє швидко та ефективно проводити пошук та аналіз правової інформації, зокрема, виявляти зв'язки між різними правовими документами, порівнювати їхні версії та встановлювати залежності між їхніми положеннями. Крім того, використання електронних конкордансів може допомогти забезпечити точність та послідовність використання термінів та понять у правових документах, що важливо для забезпечення їхньої якості та зрозумілості. В результаті, використання електронних конкордансів може значно спростити та поліпшити процес



роботи з правовою інформацією, зменшити кількість помилок та підвищити ефективність роботи правових фахівців.

В паламентському контролі електронні конкорданси можуть використовуватись для аналізу термінології та контексту в законопроектах та інших документах, які розглядаються у парламенті. Зокрема, за допомогою конкордансів можна швидко знайти всі випадки вживання певного терміна або поняття в тексті документів, що дозволяє зробити аналіз його використання та розуміння. Крім того, конкорданси можуть допомогти виявити взаємозв'язки між різними законопроектами та їх авторами, що може бути корисним для здійснення ефективного парламентського контролю. Таким чином, створення подібних корпусів є важливим науковим і практичним завданням. Створення репрезентативного корпусу є важливим національним завданням, що заохочує розвиток досліджень в області цієї мови, його використання. Якнайповнішим розміченим корпусом в Україні, наприклад, є Український національний лінгвістичний корпус (УНЛК) обсягом понад 36 млн. слів, який створено в Українському мовно-інформаційному фонді НАН України. УНЛК призначений для:

- надання текстової інформації за заданими критеріями;
- створення потоків лінгвістичної інформації для дослідницьких систем;
- інтеграції різних лінгвістично-програмних засобів у єдиному середовищі.

УНЛК як сервіс акумулює такі функціональні можливості, як сервіс електронної бібліотеки, повнотекстового пошуку, граматичного словника, лексикографічної системи «Словники України». Все це забезпечує проведення різнопланових фундаментальних лінгвістичних досліджень.

Для створення багатомовних лексикографічних систем, зокрема статистичних систем машинного перекладу застосовуються так звані паралельні корпуси текстів (Parallel Text Corpora), документів, речень, *N*-грам. Під паралельним текстовим корпусом розуміють великі зібрання паралельних текстів, тобто текстів однією мовою разом з їх перекладом іншою мовою.

На сьогодні існують алгоритми створення паралельних корпусів документів, які можна умовно розділити на дві групи: традиційні і статистичні.

До першої групи можна віднести алгоритми, за допомогою яких створювалися такі паралельні корпуси, як Корпус CRATER ([www.comp.lancs.ac.uk/linguistics/crater/corpus.htm](http://www.comp.lancs.ac.uk/linguistics/crater/corpus.htm)). Створення цих корпусів пов'язане з тим, що початкові дані завідомо паралельні.

До другої групи можна віднести паралельні корпуси, створені за допомогою статистичних алгоритмів, засновані на аналізі сторінок багатомовних веб-сайтів, заздалегідь підготовлених фрагментарних масивів і та ін.<sup>48</sup> В Україні було запропоновано новий підхід до створення паралельних корпусів документів, заснований на алгоритмі пошуку дублікатів документів, наведених різними мовами<sup>49</sup>.

До завдань корпусної лінгвістики належить і створення із вже існуючих документів спеціальних текстових корпусів, які надають інструментарій для проведення досліджень у

---

<sup>48</sup> Resnik P. Parallel strands: a preliminary investigation into mining the web for bilingual text. In D. Farwell, L. Gerber and E. Hovy (eds) *Machine Translation and the Information Soup*, Springer, Berlin, 1998. – P. 72-82.

<sup>49</sup> Жигало В.В., Ландэ Д.В. Статистический онлайн-переводчик InfoStream // Прикладна лінгвістика та лінгвістичні технології: MegaLing'2010: 36. наук. пр. / НАН України, Укр. мовно-інформ. фонд - К.: Довіра, 2010. - С. 65-78.

відносно нових для мовознавства галузях, серед яких можна назвати юридичну лінгвістику, що перебуває на межі лінгвістики і правознавства. Саме в цій галузі спостерігається значне підвищення активності.

Серед відомих міжнародних документальних корпусів з питань права можна назвати, наприклад, такі:

- Корпус LEXIS-NEXIS, що містить матеріали з більш ніж 40 країн світу.
- Корпус Westlaw, що складається з більше ніж 40 мільйонів юридичних документів з США, Великобританії та інших країн.
- Корпус HeinOnline, що містить більше 1500 видань з права та суміжних наук з різних країн світу.
- Корпус European Union Acquis, який містить всі правові акти Європейського Союзу.
- EU Parliament in XML (<http://politicalmashup.nl/eu-parliament-in-xml/>), включає усі документи Європейського Парламенту, що містяться в офіційній базі даних, з інструкціями з їх обробки).
- DutchParl. корпус парламентських документів датською мовою (<http://politicalmashup.nl/dutchparl/>), містить усі доступні у електронному вигляді парламентські документи датською мовою з Нідерланд, Фландрії і Бельгії);
- Парламентські документи з Іспанії (<http://politicalmashup.nl/spanishparliament/>), містить парламентські документи з Іспанії).

Ці корпуси включають у себе різноманітні документи, такі як закони, судові рішення, резолюції, директиви, регуляції та інші матеріали з питань права з багатьох країн світу. Вони

надають доступ до великого обсягу інформації, що сприяє більш ефективному дослідженню та аналізу правових питань.

Усталення демократії та поширення формально судових методів розв'язання конфліктів, супроводжується розширенням сфери застосування лінгвістичного аналізу та експертизи текстів документів. Засоби автоматизації лінгвістичної експертизи, що базуються на застосуванні наявних текстових корпусів підвищують ефективність семантичного аналізу текстів (процедура передбачена законодавством України) як самих законодавчих актів, так і текстів, носіями яких виступають суб'єкти та об'єкти правовідносин<sup>50</sup>.

Текстові корпуси виконують важливу роль у парламентському контролі, оскільки вони дозволяють наочно проаналізувати документи, пов'язані з роботою парламенту та його комітетів. Вони можуть бути використані для створення електронних конкордансів та підтримки парламентської діяльності в різних аспектах.

Зокрема, текстові корпуси можуть допомогти в парламентському контролі за наступними напрямками:

1. Аналіз законопроектів: з використанням корпусної лінгвістики можна зібрати велику кількість даних про терміни та поняття, що використовуються в законопроектах. Це дозволить більш об'єктивно оцінити їхню доцільність та ефективність.
2. Моніторинг діяльності депутатів: за допомогою текстових корпусів можна відстежувати, які питання та проблеми найбільше цікавлять депутатів, які є їхні основні пріоритети та дії.

---

<sup>50</sup> Широков В.А., Бугаков О.В., Грязнухіна Т.О. Корпусна Лінгвістика. – К.: Довіра, 2005. – 471 с.

3. Аналіз рішень парламенту: текстові корпуси можуть бути використані для аналізу рішень, прийнятих парламентом та його комітетами, для виявлення чіткого розуміння правових термінів та понять.
4. Підготовка звітів та аналітичних матеріалів: зібрані дані з текстових корпусів можуть бути використані для підготовки звітів та аналітичних матеріалів з різних питань, що стосуються діяльності парламенту та його комітетів.

Як можна бачити, текстові корпуси є важливим інструментом для збору та аналізу даних в контексті парламентського контролю. Вони дозволяють визначати тенденції та шляхи розвитку законодавства, аналізувати проблемні питання та потреби громадян, встановлювати зв'язки між різними законодавчими актами, виявляти негативні явища. Крім того, корпусна лінгвістика дозволяє здійснювати більш точний аналіз текстів, виявляти та аналізувати лінгвістичні особливості законодавчих текстів та їхні відмінності від звичайних текстів. В цілому, використання текстових корпусів у парламентському контролі допомагає підвищити якість законодавчих процесів, забезпечує більш точну та детальну інформацію для прийняття рішень та забезпечує більш ефективний контроль за діяльністю уряду.

#### ***Питання для самоконтролю:***

- Що таке корпусна лінгвістика і які її основні завдання?
- Які джерела використовуються для створення текстових корпусів?
- Яким чином за допомогою корпусної лінгвістики можна досліджувати правові тексти?
- Які можуть бути корисні результати застосування корпусної лінгвістики в правовій інформатиці?

- Які інші галузі можуть використовувати корпусну лінгвістику для своїх досліджень?
- Що таке текстовий корпус у лінгвістиці?
- Які розмітки можуть бути використані для текстового корпусу?
- Які завдання можуть бути вирішені за допомогою текстових корпусів?
- Які математичні методи використовуються для налаштування процесорів за допомогою текстових корпусів?
- Які проблеми можуть бути вирішені за допомогою машинного навчання на текстових корпусах?
- Що означає "збалансований склад текстів" у контексті корпусу, і як це впливає на можливості його використання?
- Яким чином корпус може бути корисним для перекладачів, і як вони можуть скористатися його функціоналом для перекладу текстів?
- Які інші можливості використання текстового лінгвістичного корпусу ви можете назвати?
- Які інші переваги можуть мати текстові корпуси у парламентському контролі окрім аналізу законопроектів?
- Яким чином використання корпусної лінгвістики допомагає об'єктивніше оцінювати доцільність та ефективність законопроектів?
- Які методи можна застосувати для моніторингу діяльності депутатів з використанням текстових корпусів?
- Які типи звітів та аналітичних матеріалів можуть бути підготовлені з використанням даних з текстових корпусів у парламентському контролі?

#### ***4.4. Семантичні мережі, онтології***

Семантична мережа – це структурована модель, що відображає зв'язки між поняттями. У семантичній мережі кожне поняття представлено вузлом, а зв'язки між ними – ребрами. Така модель дозволяє відобразити відносини між поняттями та їх значення в контексті конкретної галузі.

У правовій науці семантичні мережі застосовуються для аналізу та класифікації правових документів та понять. Наприклад, за допомогою семантичних мереж можна побудувати модель понять, що входять до складу закону, і визначити їх відносини. Також семантичні мережі можуть бути використані для розробки систем автоматичної обробки правової інформації, наприклад, систем пошуку та рекомендацій.

Існує кілька мов опису семантичних мереж, які використовуються в галузі інформаційних технологій та комп'ютерної лінгвістики. Деякі з найбільш популярних мов опису семантичних мереж включають:

1. RDF (Resource Description Framework) - це мова опису ресурсів та їх взаємозв'язків. Вона використовується для створення семантичних мереж, що базуються на відношеннях між ресурсами.
2. OWL (Web Ontology Language) - це мова опису онтологій та семантичних мереж. OWL дозволяє описувати класи, властивості та обмеження, що допомагає в створенні точних та формалізованих моделей знань.
3. SKOS (Simple Knowledge Organization System) - це мова опису схематичних засобів, яка використовується для створення та публікації тезаурусів, класифікаторів та інших схематичних засобів організації знань.

4. RDFS (RDF Schema) - це мова опису ресурсів та їх відношень, що дозволяє визначати класи, властивості та обмеження.

Ці мови опису семантичних мереж є стандартами в галузі семантичного вебу та використовуються в різних сферах, включаючи правову науку та правову інформатику.

Слід зазначити, що мережі мови та семантичні мережі – це дві різні концепції, що мають різні застосування в правовій науці.

Мережі мови – це графічне відображення зв'язків між словами чи термінами, що використовуються в текстах. Ці мережі дозволяють аналізувати взаємозв'язки між словами в текстах, що допомагає зрозуміти семантику тексту, а також виявляти певні закономірності в використанні термінів у певній галузі знань, такі як право. Мережі мови можуть відображати такі параметри, як частота вживання слів, взаємозв'язки між словами, що вказують на тематичні спільності і т.д.

Семантичні мережі, з іншого боку, використовуються для відображення семантичних зв'язків між поняттями чи термінами. Вони мають вигляд графів, де вузлами є поняття, а зв'язки між ними вказують на їх семантичні відносини. Семантичні мережі можуть використовуватись для побудови тезаурусів, систем класифікації термінів та інші цілі. В правовій науці, семантичні мережі можуть бути корисними для аналізу понять і термінів, що використовуються в законодавчих документах, та їх взаємозв'язків.

Отже, основна різниця між мережами мови та семантичними мережами полягає в тому, що мережі мови відображають взаємозв'язки між словами у текстах, а семантичні мережі - семантичні відносини між поняттями.



Одним з прикладів використання семантичних мереж у правовій науці є інтерактивний інструмент аналізу текстів Legal Concept Explorer (LCE), який дозволяє користувачам відобразити взаємозв'язки між правовими поняттями та виконувати пошук за поняттями, зв'язками між ними та відповідними правовими документами.

Legal Concept Explorer застосовує методи машинного навчання та семантичних мереж для візуалізації та аналізу правових понять та їх зв'язків.

LCE був розроблений спільно Королівським університетом в Генті та Університетом Амстердаму. Інструмент дозволяє виконувати аналіз текстів з використанням різноманітних алгоритмів машинного навчання та створювати семантичні мережі, які відображають взаємозв'язки між правовими поняттями.

LCE дозволяє виконувати пошук та аналіз текстів, що містять певне правове поняття, і відображати ці поняття у вигляді семантичної мережі, де вони пов'язані з іншими поняттями та законодавчими актами. Крім того, інструмент надає можливість порівняти різні правові поняття та знайти спільні або протилежні аспекти в їх визначенні та застосуванні.

Застосування LCE можливе в різних галузях права, включаючи кримінальне право, корпоративне право та громадянське право. Інструмент може бути використаний для аналізу законодавчих актів, судових рішень та інших документів, що містять правові терміни та поняття.

Legal Concept Explorer (LCE) має декілька важливих переваг у порівнянні з іншими інструментами для аналізу семантичних мереж в правовій науці.

По-перше, LCE є інтерактивним інструментом, що дозволяє користувачеві взаємодіяти з семантичною мережею,

використовуючи графічний інтерфейс. Це робить його більш доступним та зручним для використання, особливо для тих, хто не має глибоких знань у програмуванні та аналізі даних.

По-друге, LCE надає можливість візуалізації складних семантичних мереж, що дозволяє легко виявляти взаємозв'язки та залежності між різними концептами та термінами. Це особливо важливо для правової науки, де семантичні мережі можуть бути дуже складними та великими.

По-третє, LCE дозволяє імпортувати та аналізувати тексти з різних джерел, таких як судові рішення, законодавчі акти та інші правові документи. Це дозволяє користувачеві використовувати дані з реального світу для аналізу семантичних мереж та виявлення тенденцій у правовій науці.

Крім цього, LCE має можливість відображення концептуальних залежностей та розкриття їхнього змісту, що дозволяє більш точно та глибоко проаналізувати правові концепти та їх зв'язки.

Загалом, LCE є важливим інструментом для аналізу семантичних мереж в правовій науці, який дозволяє вивчати та розуміти правові концепти та їх взаємозв'язки, а також виявляти тенденції в галузі права.

Окрім Legal Concept Explorer, існують інші семантичні мережі, які використовуються в галузі права. Наприклад:

1. Jurisknowledgegraph - це семантична мережа, що створена для збору, організації та розуміння правової інформації. Вона містить відносини між правовими документами, юридичними поняттями та рішеннями судів.
2. UMLS (Unified Medical Language System) - це семантична мережа, розроблена для медичної галузі, але також використовується в правовій інформатиці для

аналізу термінології та взаємозв'язків між правовими поняттями.

3. SNOMED CT (Systematized Nomenclature of Medicine - Clinical Terms) - це семантична мережа, що використовується в медичній термінології, але також має застосування в правовій інформатиці для створення та управління термінологією та концептуальними моделями.
4. EuroVoc - це тезаурус, що розроблений Європейським Союзом для категоризації та структурування офіційної інформації. Він містить понад 7 тисяч термінів, що пов'язані з законодавством ЄС.
5. WordNet - це семантична мережа, що розроблена для англійської мови, але має застосування в галузі права для аналізу синонімів та взаємозв'язків між правовими термінами.

Окрема слід зупинитись на комп'ютерних онтологіях у галузі права.

Комп'ютерна онтологія - це формальна специфікація концептуальної моделі, яка описує концепти та взаємозв'язки між ними в певній галузі знань або домені. Онтології допомагають зрозуміти смисл термінів та понять, що використовуються в певній галузі, і можуть бути використані для автоматичної обробки текстової інформації та пошуку даних.

Термін «онтологія» використовується в декількох областях знань і має два різних значення:

- філософська дисципліна, яка вивчає найбільш загальні характеристики буття і сутностей;
- в інженерії знань: артефакт, структура, модель знань, що описує значення елементів деякої системи.

У філософії онтологією називають теорію про природу буття і види сутностей. В інженерії знань онтологічний рівень формалізує накопичені знання, визначаючи і поєднуючи термінологію різних предметних сфер. Таким чином, чіткої взаємної обумовленості між значеннями терміна «онтологія» у філософії і в інженерії знань не простежується. Зв'язок між ними носить скоріше асоціативний характер.

Незважаючи на існування великої кількості напрацювань у галузі представлення знань, не існує єдиного чіткого визначення онтологій.

Під комп'ютерною онтологією в рамках цієї роботи розумітимемо систему понять предметної галузі, яка представляється як набір сутностей, що об'єднані різними відношеннями.

Одна з головних відмінностей між онтологіями та семантичними мережами полягає у тому, що онтології дозволяють використовувати формальні, строго визначені та структуровані описи понять та їх взаємозв'язків. У той час, як семантичні мережі можуть бути менш формалізованими та більш зорієнтованими на візуальне представлення та розуміння понять із контексту.

Ще одна відмінність полягає у тому, що онтології зазвичай використовуються для автоматичної обробки та аналізу даних, в той час як семантичні мережі - для візуалізації та розуміння взаємозв'язків між поняттями.

Онтології отримують досить широке поширення в задачах представлення знань, семантичної інтеграції інформаційних ресурсів, інформаційного пошуку тощо. У науці про «штучний інтелект» онтологія – це «специфікація концептуалізації предметної області», або спрощено, документ або файл, що формально задає зв'язки між поняттями. Це свого роду словник понять предметної сфери і сукупність явно визначених припущень щодо змісту цих

понять. Найчастіше онтологія представляється як ієрархія понять, пов'язаних відношеннями певних видів. Розвинуті онтології формалізуються засобами мов логіки і допускають можливості формування логічних тверджень.

Онтології використовуються для формальної специфікації понять і відношень, які характеризують певну область знань. Перевагою онтологій як способу представлення знань є їх формальна структура, яка спрощує їх комп'ютерну обробку.

Терміну «онтологія» задовольняє широкий спектр структур, що представляють знання про ту чи іншої предметної області. Так до онтології можна віднести ряд структур, що відрізняються різним мірою формалізації:

- глосарій;
- проста таксономія;
- тезаурус (таксономія з термінами);
- понятійна структура з довільним набором відношень;
- повністю аксіоматизована теорія.

У загальному вигляді структура онтології являє собою набір елементів чотирьох категорій:

- поняття;
- відношення;
- аксіоми;
- окремі екземпляри.

Дослідники виділяють прикладні онтології, онтології області знання, загальні (родові) онтології і репрезентаційні онтології (йдеться щодо онтологій метарівня, що включають в себе репрезентаційні першоелементи).

Онтології можуть бути також розділені на одномовні і багатомовні. Вже існує ряд онтологій, орієнтованих на

представлення знань на декількох мовах, наприклад, EuroWordNet, MikroKosmos і деякі інші.

Також виділяється особливий тип онтологій – лексичні онтології (або лінгвістичні). Відмітною властивістю таких онтологій є «фіксація в одному ресурсі понять (слів) разом з їх мовними властивостями». Такі онтології тісно взаємопов'язані з семантикою граматичних елементів (слів, іменних груп та ін.) Основним джерелом понять в онтологіях цього типу є значення мовних одиниць. Їх також відрізняє своєрідний набір відношень, зазвичай властивий для мовних елементів: синонімія, гіпонімія, меронімія, а також ряд інших. До лінгвістичних онтологій відносяться WordNet, MikroKosmos, Sensus, RuГез та інші. Коло завдань, що вирішуються такими онтологіями, тісно взаємопов'язане з обробкою природної мови.

Онтології, зокрема, можна ефективно використовувати для підвищення точності інформаційного пошуку – пошукова система буде видавати тільки такі документи, де потрібне поняття згадується точно, а не ті, у текстах яких зустрілося задане ключове слово.

У галузі права, онтології використовуються для формалізації термінології та взаємозв'язків між різними поняттями, що використовуються в законодавстві та юридичній практиці. Онтології можуть допомогти у вирішенні таких задач, як автоматичний пошук документів, аналіз текстів, розуміння контексту і виявлення залежностей між різними поняттями.

Одна з головних відмінностей між семантичними мережами та онтологіями полягає в тому, що останні мають більш формальний та структурований підхід до опису понять та взаємозв'язків між ними. Онтології зазвичай базуються на математичних моделях та формальних правилах, що дозволяє

забезпечити більш точне та структуроване опис понять та взаємозв'язків між ними. Семантичні мережі, з іншого боку, можуть бути менш формалізовані та більш зорієнтовані на графічне представлення взаємозв'язків між поняттями.

Прикладами онтологій у галузі права є LegalRuleML, SWRL, LKIF, Protege-OWL та інші. Онтології у галузі права є важливим інструментом для створення формалізованих описів правових понять, відносин між ними та правових актів. Наприклад, LegalRuleML - це мова високого рівня для формалізації правил, які використовуються в правових системах. SWRL (Semantic Web Rule Language) - це мова правил, яка базується на OWL (Web Ontology Language) та RDF (Resource Description Framework), що дозволяє використовувати онтології у веб-програмуванні. LKIF (Legal Knowledge Interchange Format) - це мова, яка використовується для формалізації правових знань та представлення їх у машинно-читабельному вигляді. Protege-OWL - це інструмент для розробки онтологій, який дозволяє створювати формалізовані описи для понять та відносин між ними в галузі права.

Protege-OWL (Ontology Web Language) є однією з найпопулярніших мов опису онтологій у галузі права та інших галузях. Це безкоштовний інструмент для розробки онтологій, який дозволяє створювати, редагувати та розширювати онтології з використанням графічного інтерфейсу.

Protege-OWL дозволяє користувачам створювати онтології з використанням стандартних понять та відносин, які можуть бути використані для опису певної галузі знань, включаючи галузь права. Онтології, створені за допомогою Protege-OWL, можуть бути інтегровані з іншими онтологіями та базами знань, що дозволяє створювати комплексні системи знань.

Protege-OWL також має низку додаткових функцій, які дозволяють користувачам тестувати та перевіряти онтології на відповідність стандартам OWL і RDF. Крім того, Protege-OWL може бути інтегрований з іншими інструментами, такими як системи автоматичного розпізнавання мови та системи пошуку інформації.

Загалом, Protege-OWL є потужним інструментом для створення онтологій та розробки комплексних систем знань у галузі права та інших галузях. Він дозволяє легко створювати та редагувати онтології, тестувати та перевіряти їх на відповідність стандартам, інтегрувати їх з іншими інструментами та створювати комплексні системи знань.

Заслуговує на увагу також проект Estrella ([www.estrellaproject.org](http://www.estrellaproject.org), Рис. 3), в рамках якого розроблено онтологію LKIF (Legal Knowledge Interchange Format) – мову для подання юридичних знань та обміну між правовими інформаційними базами.

На даний момент для розробки систем, заснованих на знаннях, є актуальною задача об'єднання різних репрезентативних підходів з метою забезпечення найбільш повного подання знань у правовій сфері. У рамках розробки базової правової онтології LKIF-Rus на базі міжнародної онтології LKIF-Core (LKIF – Legal Knowledge Interchange Format) було створено онтологію цивільного права LKIF-CivilRus.



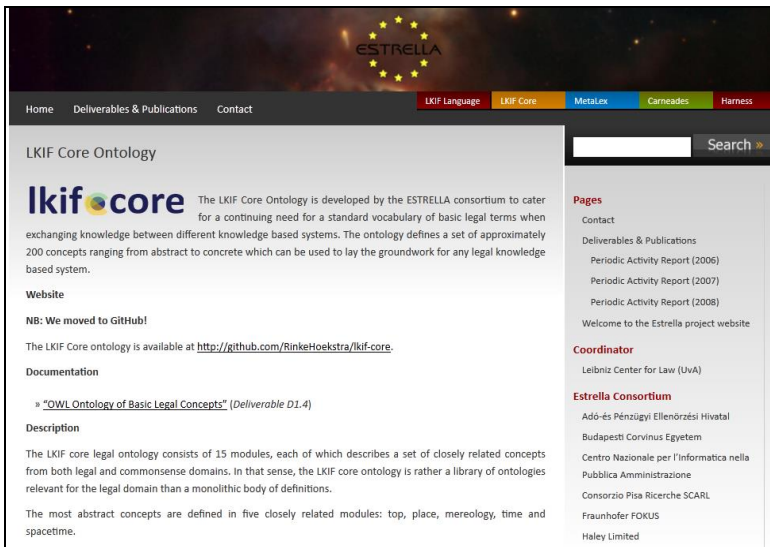


Рисунок 3 – Сторінка LKIF на веб-сайті проекту Estrella

Сучасні засоби онтологічного моделювання дозволяють частково впровадити продуктивний підхід в процес розробки онтології. Для цього, наприклад, можна використовувати SWRL-правила (SWRL – Semantic Web Rule Language), підтримка яких включена в середовище розробки Protégé.

Онтології є корисним інструментом для задач парламентського контролю, оскільки вони дозволяють описати концептуальну структуру домену, що досліджується. Онтології можуть використовуватися для розуміння зв'язків між різними аспектами досліджуваного домену, а також для організації та структурування даних, пов'язаних з доменом.

У контексті парламентського контролю, онтології можуть допомогти виконанню таких завдань, як збір та обробка інформації про законопроекти, розробка системи підтримки прийняття рішень для депутатів, аналіз різних статистичних даних, що стосуються діяльності парламенту, та багато іншого.

Одним з прикладів використання онтологій в парламентському контролі є проект "LinkedPolitics", який розробляється Європейським парламентом. Цей проект використовує онтології для створення системи, яка забезпечує інформаційну підтримку для депутатів та допомагає у вирішенні різних завдань, пов'язаних з роботою у парламенті.

Існують деякі особливості використання онтологій для подання юридичних знань:

1. По мірі розвитку будь-якої правової системи в нормативні акти вводяться нові або видаляються попередні причинно-наслідкові зв'язки, що може призвести до необхідності перевизначення термінів, зміни їх положення в таксономії. Таким чином онтологію необхідно постійно змінювати. У зв'язку з цим в онтологічному моделюванні є цілий напрямок – управління версіями (versioning).
2. Розробник онтології не може гарантувати, що визначення повністю відобразить сенс юридичного поняття (принаймні, якщо це визначення не з нормативного акту).
3. Не завжди можливим є вираз причинно-наслідкових сутностей правових явищ.
4. Протиріччя між вимогою однозначного визначення термінів у рамках правової інформації та практикою призводить до неузгодженості онтології, що неприпустимо.

Серед проблем, які виникають в процесі розробки міжнародних правових онтологій, найбільш суттєвими можна назвати наступні:

- 1) відмінності в правових системах і, як наслідок, у правовому понятійному апараті;
- 2) багатозначність деяких термінів, синонімія;
- 3) проблеми при позначенні відношень і особливо зворотних відношень.

### ***Питання для самоконтролю***

- Що таке семантична мережа та як вона відображає зв'язки між поняттями?
- Які можливості надає застосування семантичних мереж у правовій науці?
- Які мови опису семантичних мереж використовуються в галузі інформаційних технологій та комп'ютерної лінгвістики?
- Яку роль відіграє мова OWL в створенні точних та формалізованих моделей знань?
- Яку функцію виконує мова SKOS та для чого вона використовується?
- Що таке мережі мови і як вони допомагають аналізувати текст?
- Яка основна різниця між мережами мови та семантичними мережами?
- Для чого використовують семантичні мережі в правовій науці?
- Який інтерактивний інструмент використовують для аналізу текстів у правовій науці?
- Які параметри можуть бути відображені в мережах мови?
- Що розуміється під комп'ютерною онтологією та як вона представляється?

- Яка відмінність між онтологіями та семантичними мережами?
- Які можливості надає онтологія в задачах представлення знань та інформаційного пошуку?
- Яким чином розвинуті онтології формалізуються засобами мов логіки?
- Яким чином семантичні мережі відрізняються від онтологій у візуальному представленні понять?
- Які можливості надають онтології для задач парламентського контролю?
- Яким чином онтології допомагають організувати та структурувати дані, пов'язані з досліджуваним доменом?
- Які завдання у контексті парламентського контролю можуть виконуватися з використанням онтологій?
- Які переваги має використання онтологій для аналізу статистичних даних, що стосуються діяльності парламенту?
- Яким чином онтології допомагають розуміти зв'язки між різними аспектами досліджуваного домену?

## **5. ПОРІВНЯЛЬНИЙ АНАЛІЗ ДОКУМЕНТІВ**

Порівняльний аналіз документів є важливим завданням у праві та парламентському контролі, яке полягає у зіставленні текстів законів, рішень судів, договорів та інших документів для виявлення спільних та відмінних рис. Це дозволяє отримати більш повне розуміння правової системи, знайти підстави для судових рішень та сприяє покращенню якості законодавства.

Порівняльний аналіз може здійснюватися вручну або за допомогою спеціальних програм для обробки текстів. Останні дозволяють автоматично порівнювати тексти документів та виділяти їх схожі та відмінні риси. Такі програми використовують різні методи порівняння, наприклад, порівняння за ключовими словами, порівняння за синтаксичною структурою, порівняння за семантикою та інші.

У парламентському контролі порівняльний аналіз документів може бути корисним для виявлення змін у законодавстві, знаходження відмінностей між версіями законів, аналізу рішень судів та інших документів. Завдяки цьому можна забезпечити більш ефективний контроль за виконанням законів та забезпечити більш якісну правову підтримку діяльності парламенту.

У праві порівняльний аналіз документів є важливим інструментом для забезпечення консистентності законодавства (відповідність законів і правових актів одне одному, а також відсутність суперечностей і конфліктів між ними.) та виявлення різних правових підходів до одних і тих самих питань. Наприклад, порівняння законодавства різних країн дозволяє зрозуміти відмінності між правовими системами та знайти найкращі практики для покращення власної системи.

Порівняння документів також може допомогти у виявленні різниці між законами та їх тлумаченням в різних країнах або регіонах. Це може допомогти законодавцям зробити більш обґрунтований вибір при створенні нових законів.

У парламентському контролі порівняльний аналіз документів може допомогти виявити різні правові підходи до одних і тих самих питань, що може бути корисно при прийнятті рішень та розробці нових законопроектів. Також, порівняння документів може допомогти у виявленні ризиків та прогалин у законодавстві, що дозволить вчасно вжити необхідних заходів.

Отже, порівняльний аналіз документів є важливим інструментом у праві та парламентському контролі, що дозволяє виявляти різні правові підходи, проблеми та прогалини в законодавстві, та забезпечувати більш обґрунтоване прийняття рішень.

Проблема порівняльного аналізу електронних текстів постала практично одночасно з появою можливостей обробки текстів комп'ютерною технікою. Друга половина ХХ-го століття характеризувалася становленням цього напрямку, що обумовлюється бурхливим розвитком формальної і комп'ютерної лінгвістики.

Разом з тим, слід відзначити, що фундаментальні роботи у галузі порівняння текстів проводилися ще значно раніше, наприклад, видатна робота з цього приводу була надрукована ще у 1915 р.<sup>51</sup>

На даний час з розвитком мережі Інтернет задачі порівняльного аналізу електронних текстів мають

---

<sup>51</sup> Морозов Н.А. Лингвистические спектры: средство для отличения плагиатов от истинных произведений того или иного известного автора. Стилометрический этюд. // Известия отд. русского языка и словестности Импер. Акад. наук, Т. XX, кн. 4, 1915.

відношення до ряду таких технологічних напрямків, як інформаційний пошук, узагальнення та групування інформації. Розвиток Інтернету також визвав значне зростання дубльованої та запозиченої інформації у різних сферах: освіти, засобах масової інформації, іноді в науці, в практиці законотворчості, що пов'язано з проблемою гармонізації вітчизняного і міжнародного законодавства.

На цей час визначено п'ять таких напрямків порівняльного аналізу електронних текстів<sup>52</sup>:

- аналіз унікальності документів інформаційно-пошуковими системами в Інтернеті для запобігання зайвої індексації однакових документів;
- виявлення передруку, випадків плагіату, перевірка коректності запозичень у нормативно-правових документах;
- архівування документів (зменшення обсягів даних на комп'ютерних носіях);
- кластеризація документів за їх подібністю, виявлення кластерів близьких за змістом документів, виявлення основних тематик в інформаційних потоках.
- пошук та фільтрація спаму.

Динаміка інформаційних потоків у веб-просторі, соціальних та пірінгових мережах, обумовлює і ряд проблем, що виникають при порівнянні електронних текстів. При порівнянні електронних текстів можуть виникати різні проблеми, такі як:

---

<sup>52</sup> Osipovs P., Borisov A. Practice of Web Data Mining Methods Application // J. Riga Technical University 40: 101-107 (2009) . – P. 11-18.

1. Форматування тексту: різні програми для обробки тексту можуть застосовувати різне форматування, що може ускладнити порівняння документів.
2. Наявність помилок та опечаток: навіть найточніші системи розпізнавання тексту можуть допускати помилки, що може призвести до неточностей у порівнянні.
3. Різні мови та алфавіти: якщо порівнюються документи, написані різними мовами або використовуються різні алфавіти, то можуть виникати складнощі зі збігом та порівнянням окремих символів.
4. Змістовна різниця: дві версії документа можуть мати різний зміст, навіть якщо текст формально однаковий. Це може статися, наприклад, якщо в одній версії документа було вилучено деякі розділи або внесено додаткові зміни.
5. Об'єм документів: порівняння великої кількості документів може бути дуже часо- та ресурсомістким завданням, що потребує використання спеціальних програм та алгоритмів.

Задача виявлення нечітких дублікатів є однією з актуальніших і складніших, особливе практичне значення вона приймає, зокрема, при інтеграції інформаційних ресурсів, боротьбі з плагіатом, визначенні спам-розсилок тощо. Ця задача полягає в ідентифікації подібних або ідентичних документів, які мають деякі відмінності, такі як різні формати файлів, різна орфографія та граматики, зміна форматування та інші несуттєві зміни у тексті. Це завдання дуже важливе в багатьох галузях, зокрема у правовій сфері, де велика кількість документів містить схожі висновки, але можуть містити різні формулювання.



Один з підходів до виявлення нечітких дублікатів полягає в застосуванні алгоритмів обробки природних мов, таких як алгоритм Левенштейна, який використовується для порівняння послідовностей символів і визначення відстані між ними. Інші підходи включають використання хеш-функцій для порівняння документів та застосування методів машинного навчання, таких як нейронні мережі, для виявлення схожих текстів. Важливо враховувати, що нечіткі дублікати можуть мати різні рівні подібності, і відповідно, необхідні різні методи для їх виявлення. Також важливо враховувати, що автоматичне виявлення нечітких дублікатів може бути не повністю точним, тому їх результати повинні бути перевірені вручну фахівцем перед їх використанням у правовій практиці.

У багатьох випадках в основу пошуку нечітких дублікатів, зокрема, плагіату покладено порівняння текстів<sup>53</sup>. Спочатку проводиться порівняння текстів в цілому, а далі відбувається розбивка на абзаци і потім пошук конкретних фрагментів тексту в інших документах. В інших же випадках використовується пошук за ключовими словами або ж словосполученнями. Враховуючи кількість термінів з малою частотою, знайдених при перевірці, як результат отримують підтвердження того, що документ або фрагмент є плагіатом, або навпаки. Багато систем виконують пошук не тільки подібних фрагментів, але й проводять достатньо складний аналіз тексту, що включає, наприклад, використання методу Флешу, який дозволяє обчислити індекс «легкості» тексту. Аналізуючи показники індексів абзацив у роботі, що перевіряється, можна ідентифікувати аномалії, найбільш ймовірні абзаци плагіату, пошук яких потім проводиться у внутрішній базі даних, або ж у веб-просторі.

---

<sup>53</sup> Stone W.R. Plagiarism, Duplicate Publication and Duplicate Submission: They Are All Wrong! // IEEE Antennas and Propagation, Aug. 2003. – 45. – № 4. – P. 47-49.

Серйозне спрощення названої задачі може бути отримано за рахунок застосування формальних методів, наприклад, математичної статистики, сигнатурних алгоритмів, репутаційних підходів (наприклад, шляхом ранжирування першоджерел, тематик, опорних слів тощо).

На формальному рівні, шляхом зіставлення фрагментів текстів, шинглів, лінгвістичних сигнатур тощо, сьогодні з успіхом виявляються нечіткі дублікати, що формуються шляхом копіювання, прямого перекладу з іноземних мов, компіляції тощо.

Однак нечіткі дублікати включають також результати переробки оригіналів на змістовному рівні, наприклад, перекази тих самих подій, текстів, ідей, опис різних аспектів різними авторами. Крім того, нечіткі дублікати можуть бути представлені в різних медіа-середовищах. В даний момент подібність узагальнених таким чином документів не завжди може бути визначена за допомогою нижченаведених методів та алгоритмів. Саме тому порівняльний аналіз електронних текстів являє собою нині відкриту науково-практичну проблему.

### ***5.1. Проблематика порівняльного аналізу документів***

Якщо деякий документ повністю збігається за текстом з іншим документом, то кажуть що має місце чіткий дублікат. Якщо документ за текстом співпадає не повною мірою, але є збіг за змістом, то кажуть про «майже дублікат», «нечіткий дублікат», або «текстовий синонім»<sup>54,55</sup>. Текстовий синонім є розширенням поняття синонімії звичайних слів або

---

<sup>54</sup> Bourdaillet J. Alignment of Noisy Unstructured Text Data // IJCAI-2007 Workshop on Analytics for Noisy Unstructured Text Data. Hyderabad, India – January 8, 2007. – P. 139-146.

<sup>55</sup> Широков В.А. Елементи лексикографії.– К.: Довіра, 2005. – 304 с.

словосполучень до повних текстів, які мають властивості семантичної подібності. Виходячи з цього якісного розуміння у наведеній роботі було надано визначення: «назвемо текст  $T_1$  слабко синонімічним текстові  $T_2$ , якщо ці тексти відрізняються за формою, тобто за формальним зовнішнім виглядом, але близьким за змістом».

Запобігання використанню інформації, що явно дублюється, не становить проблем, проте подібні за змістом документи знаходяться не так легко. На практиці явні дублікати виявляються за допомогою так званих сигнатурних механізмів: контрольних сум, хешів, але цей підхід не вирішує всіх проблем користувачів, для яких частіше за все не має значення, з чим вони мають справу, з прямим передруком або з перефразуванням.

Загальну традиційну схему виявлення нечітких дублікатів документів наведено на Рис. 4.

Важливою вимогою при реалізації алгоритмів виявлення нечітких дублікатів документів є їхня стійкість по відношенню до невеликих змін змісту документів, що аналізуються. При цьому слід відзначити, що в технологічному ланцюжку загальної схеми виявлення дублікатів документів застосовується обчислення контрольних сум, які були розроблені для задач практичної криптографії. Для виявлення дублікатів документів можна використовувати контрольні суми і хеш-функції. Контрольна сума - це значення, яке визначається для певного документа на основі його вмісту. Це може бути, наприклад, MD5-хеш, SHA-1-хеш або SHA-256-хеш. Якщо два документи мають однакову контрольну суму, то це може свідчити про те, що вони є дублікатами.

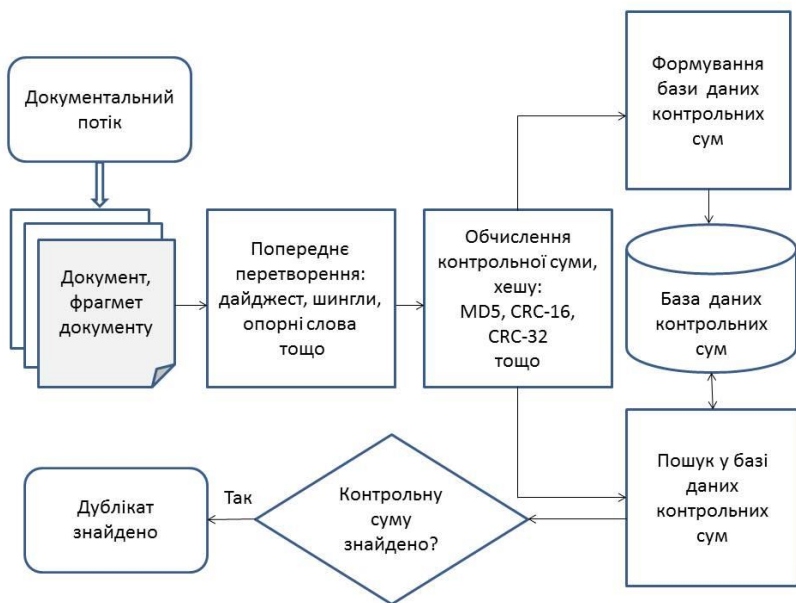


Рисунок 4 – Загальна схема технології виявлення нечітких дублікатів

Хеш-функція – це алгоритм, який приймає на вхід будь-який великий блок даних і повертає унікальне фіксоване значення (хеш) фіксованої довжини. Хеш-функції також можуть використовуватися для виявлення дублікатів документів. Проте, вони можуть бути менш надійними в порівнянні з контрольними сумами, оскільки може бути певна ймовірність, що два різних документи матимуть однаковий хеш.

Крім того, можна використовувати й інші методи для виявлення дублікатів документів, такі як порівняння текстів, використання алгоритмів машинного навчання, які базуються на векторному поданні текстів тощо.

Необхідно відзначити, що відповідні алгоритми орієнтовані на прямо протилежну задачу, а саме, розсіювання

інформації, тобто невеликі зміни у змісті вихідних документів мають призводити до кардинальних змін у контрольних сумах, хешах. Саме це протиріччя, є однією з передумов виникнення помилок, які зустрічаються у сучасних системах виявлення нечітких дублікатів.

У якості найбільш вживаних сучасних алгоритмів порівняння строкових даних можна назвати метод Карпа-Рабіна<sup>56</sup> і сигнатурні методи, у тому числі метод шинглів, запропонований А. Брьодером<sup>57</sup> та його розвиток – метод супершинглів. Велике практичне значення мають методи лексичних сигнатур, які враховують мовну природу документів, що порівнюються.

Методи, засновані на урахуванні повторень ланцюжків слів, наприклад, метод «шинглів», докладно описано в роботах<sup>58,59</sup>. Однак ці ефективні в багатьох випадках методи пошуку «майже дублів» виявилися не дуже чутливими для невеликих правових текстів з великим обсягом перефразувань.

Сьогодні стало звичайним звернення до статистичних підходів. Відома методика виявлення дублікатів, засновану на аналізі  $N$  найбільш «якісних» слів. При цьому якість слів визначалася експертами, а відповідний математичний апарат одержав назву «нечіткої цифрової сигнатури». При цьому використовується так званий «наївний підхід» (множення ймовірностей залежних подій – слів у повідомленнях), а

---

<sup>56</sup> Karp R.M., Rabin M.O. Efficient randomized pattern-matching algorithms // IBM Journal of Research and Development. – 31 (2), March 1987. 249-260.

<sup>57</sup> Broder A. Identifying and Filtering Near-Duplicate Documents, COM'00 // Proceedings of the 11th Annual Symposium on Combinatorial Pattern Matching. – 2000. – P. 1-10.

<sup>58</sup> Manber U. Finding similar files in a large file system // Proceedings of the 1994 USENIX Conference, January 1994. – P. 1-10.

<sup>59</sup> Broder A., Glassman S.C., Manasse M.S. Syntactic Clustering of the Web // WWW6, 1997.

також елементи «ручного» відбору значимих слів (очевидно, важливість окремих слів може змінюватися у часі).

Першим кроком формалізації відношення подібності або слабкої синонімії (умовно будемо вважати – семантичної близькості, але, зрозуміло, що друге поняття – більш загальне) є введення відповідної функції  $F(X, Y)$ , яка ставить у відповідність деякій парі документів  $(X, Y)$  деяке дійсне число. Функцію  $F(X, Y)$  визначено в околі  $[0, 1]$ , тобто  $0 \leq F(X, Y) \leq 1$ . Необхідною і достатньою умовою співпадіння  $X$  та  $Y$  є  $F(X, Y) = 1$ .

Важлива якість відношення подібності – несиметричність, тобто, у загальному випадку:

$$F(X, Y) \neq F(Y, X).$$

У подальшому викладенні перейдемо до скорочених позначень, а саме  $F(X, Y) > 0$  позначимо як  $X < Y$  (" $<$ " – відношення подібності), а  $F(X, Y) = 1$  позначимо як  $X \equiv Y$  (" $\equiv$ " – відношення точного дублювання).

Очевидно, що для відношення подібності і точного дублювання справедливі правила рефлексивності:

$$A < A, \quad A \equiv A,$$

де  $A$  – довільний документ.

Відношення подібності не має властивості симетричності. Із подібності документа  $A$  документу  $B$  не випливає зворотне, тобто:

$$A < B \not\Rightarrow B < A.$$

Також не виконується умова транзитивності:

$$A < B, \quad B < C \not\Rightarrow A < C.$$

Дійсно, наприклад, окремий документ може бути подібний до тексту з добірки, яка його включає, але сама добірка може не бути подібною до цього документа. Або документ може бути подібний до двох документів, з яких він скопійований, але самі оригінали можуть суттєво відрізнятись від нього.

Для відношення дублювання, навпаки, симетричність і транзитивність виконуються:

$$A \equiv B \Rightarrow B \equiv A,$$

$$A \equiv B, B \equiv C \Rightarrow A \equiv C.$$

Зауважимо, що відношення, яке має властивості рефлексивності, симетричності і транзитивності є відношенням еквівалентності, у нашому випадку, відношенням змістовного збігу або дублювання.

Як було відзначено, властивість дублювання документів є більш жорстким критерієм подібності, наприклад, збіг 3, 4 або 5 термів свідчить про деяку змістовну близькість, тобто можна записати:

$$" < " \Rightarrow " \equiv ".$$

#### ***Питання для самоконтролю:***

- Які є різниці між чітким та нечітким дублікатом документу?
- Які методи використовуються для виявлення явних дублікатів документів?
- Як можна визначити нечіткий дублікат документу?
- Яка важлива вимога до алгоритмів виявлення нечітких дублікатів документів?
- Які застосовуються методи обчислення контрольних сум та хеш-функцій для виявлення дублікатів документів?
- Які методи порівняння строкових даних є найбільш вживаними сьогодні?

- Яку методику виявлення дублікатів використовують, що базується на аналізі  $N$  найбільш "якісних" слів?
- Як визначається якість слів у цій методиці та який математичний апарат використовується?
- Які проблеми виникають при використанні методів порівняння строкових даних для невеликих правових текстів з великим обсягом перефразувань?

## 5.2. Алгоритми виявлення подібних документів

### Алгоритм Карпа-Рабіна

Алгоритм Карпа-Рабіна є алгоритмом, який використовується для пошуку підрядка у тексті. Він базується на хешуванні підрядка та порівнянні його з хешами підрядків тексту. Введемо деякі позначення. Очевидно, будь-який рядок у пам'яті комп'ютера представляється послідовністю байтів, кожний з яких є послідовністю бітів, тобто двійкових значень. Позначимо  $T$  – двійковий рядок довжиною  $|T| = m$ ;  $P$  – двійковий шаблон для пошуку довжиною  $|P| = n$ . Введемо функцію:

$$H(P) = \sum_{i=0}^{n-1} 2^{n-i} P(i),$$

де  $P(i)$  –  $i$ -й біт рядку  $P$ .

Також визначається функція від підрядка  $T$  довжиною  $n$  ( $T_r^n$ ), яка починається з  $r$ -го біту:

$$H(T_r^n) = \sum_{i=0}^{n-1} 2^{r+n-i} T_r^n(r+i-1).$$

З того, що будь-яке ціле число можна єдиним чином представити у вигляді суми позитивних степенів двійки,



впливає, що підрядок  $T_r^n$  входить у  $T$  починаючи з позиції  $r$  у тому й тільки у тому разі, коли  $H(P) = H(T_r^n)$ .

При пошуку підрядку у рядку, таким чином, при порівнянні  $H(P)$  і  $H(T_r^n)$ , з того, що представлення  $2^n$  числа вимагає  $n$  бітів, впливає, що необхідні для порівняння числа експоненційно великі, тобто задача порівняння виявляється експоненційно складною, тобто практично не здійсненою.

У 1987 році Р. Карп і М. Рабін оприлюднили алгоритм, який отримав назву методу рандомізованих дактилограм, в якому суттєво знижена складність обчислень. Але цей алгоритм дозволяє стверджувати про входження підрядку у рядок не абсолютно точно, а з деякою високою ймовірністю. Надамо стислий зміст цього алгоритму.

Нехай  $H_p(P) = H(P) \bmod p$  – залишок від ділення  $H(P)$  на  $p$ . У випадку, коли  $p$  – просте число ( $p \ll n$ ), залишок від ділення  $H_p(P)$  і  $H_p(T_r^n)$  на  $p$  називають дактилограмами  $P$  і  $T_r^n$  по модулю  $p$ . Звісно,  $0 \leq H_p(P), H_p(T_r^n) \leq p-1$ .

Якщо  $P$  входить до  $T$ , починаючи з позиції  $r$ , то  $H_p(P) = H_p(T_r^n)$ , але зворотне, не вірно. Кажуть, що коли  $H_p(P) = H_p(T_r^n)$ , але  $P$  не входить до  $T$ , то має місце помилковий збіг  $P$  і  $T_r^n$  з позиції  $r$ . Для оцінки ймовірності відсутності помилкових збігів введемо позначення  $\pi(q)$  – кількість простих чисел, що не перевершують  $q$ . Доведено, що має місце наступна нерівність:

$$\frac{q}{\ln q} \leq \pi(q) \leq 1.26 \frac{q}{\ln q}.$$

Крім того, має місце теорема, що справедлива для  $P$  і  $T$ , при умові  $nm \geq 29$  ( $n = |P|$ ,  $m = |T|$ ). Якщо  $I$  – будь яке позитивне число, а  $p$  – випадковим чином вибране просте

число, що не перевищує  $I$ , то ймовірність помилкового збігу  $P$  і  $T$  не перевищує  $\pi(nm)/\pi(I)$ .

Звідси впливає такий алгоритм випадкової дактилограми для пошуку входжень  $P$  до  $T$ :

1. Вибрати позитивне ціле число  $I$ .
2. Випадковим чином вибрати просте число  $p$ , що не перевершує  $I$ , та обчислити  $H_p(P)$ .
3. Для кожної позиції  $r$  у  $p$  обчислити  $H_p(T_r^n)$  і зіставити з  $H_p(P)$ . Якщо вони рівні, то або оголосити про ймовірний збіг, або у явному вигляді перевірити збіг  $P$  з  $T$ , починаючи з позиції  $r$ .

Алгоритм Карпа-Рабіна має складність  $O(n + m)$ , де  $n$  – довжина тексту, а  $m$  – довжина підрядка. В порівнянні з іншими алгоритмами пошуку підрядків, він є досить швидким, особливо якщо маємо справу з великими текстами. Алгоритм Карпа-Рабіна може бути використаний для виявлення дублікатів документів шляхом порівняння хешів текстів. Однак, він може давати помилкові співпадіння у випадках, коли два текста мають різний зміст, але мають однакові хеші. Тому, для більш точного виявлення дублікатів, можуть використовуватись інші методи, такі як порівняння за змістом та структурні методи.

### ***Алгоритм «шинглів»***

Через десять років після оприлюднення алгоритму Кара-Рабіна А. Брьодер і його співавтори<sup>60</sup> представили свій алгоритм, що базується на оцінці подібності документів за збігом послідовностей з визначеної кількості  $n$  сусідніх слів.

---

<sup>60</sup> Broder A., Glassman S.C., Manasse M.S. Syntactic Clustering of the Beб // WWW6, 1997.

Такі послідовності автори назвали шинглами (від англ. *Shingles* – «лусочки»). Необхідно зазначити, що різні шингли формуються з послідовностей слів в нахліст, а не впритул, тобто наступний шингл починається з наступного слова, а не із слова з номером, більшим на довжину шинглу. Два документи вважаються дублікатами, якщо множини їх шинглів суттєво перетинаються.

Таким чином, розбиваючи текст на послідовності слів (їх ще називають  $N$ -грамами), ми отримуємо набір шинглів у кількості  $N - n + 1$ , де  $N$  – кількість слів у документі.

При цьому на шингли розбивається кожний екземпляр документів, що порівнюються. Оскільки кількість шинглів, відповідних кожному документу  $\epsilon$ , таким чином, досить великою, було запропоновано декілька методів їх зменшення для отримання репрезентативних підмножин для порівняння.

Перший запропонований метод полягав у тому, що розглядалися лише ті шингли, чиї дактилограми, що обчислювалися за алгоритмом Карпа-Рабіна, ділилися без залишку на деяке число  $m$ . Основний недолік цього підходу – залежність вибірки від довжини документа, і тому невеликим за розміром документам відповідають дуже короткі вибірки, що призводить до зменшення якості виявлення дублікатів.

У відповідності з іншим запропонованим методом відбиралася лише фіксована кількість  $S$  шинглів з найменшими значеннями дактилограм або залишалися всі шингли, якщо їх загальна кількість не перевищувала  $S$ .

Подальшим розвитком методу шинглів  $\epsilon$  методи «супершинглів» та «мегашиноглів»<sup>61</sup>.

---

<sup>61</sup> Broder A. Identifying and Filtering Near-Duplicate Documents, COM'00 //

Метод супершинглів полягає у тому, що для кожного документа вибираються випадкові набори шинглів у кількості 84. Для кожного вибраного шинглу підраховується дактилограма Карпа-Рабіна. Після цього 84 шингли розбиваються на 6 груп (супершинглів) по 14 шинглів у кожній. У результаті кожний документ представляється 6 супершинглами. Виявляється, що при умові подібності документів на рівні 95%, ймовірність збігу 2-х супершинглів становить приблизно 90%, а якщо подібність між документами лише 80%, то ймовірність збігу щонайменш двох шинглів становить лише 2,6%. Таким чином, для ефективного порівняння документів виявляється достатнім дослідити збіг лише однієї пари шинглів.

Ідея об'єднання шинглів отримала подальший розвиток в методі мегашинглів, який полягає у тому, що для кожного документа розглядаються усі пари його супершинглів. Кількість таких пар дорівнює  $C_6^2 = \frac{6 \cdot 5}{2} = 15$ . Стверджується, що два документи є щонайменше нечіткими дублікатами, якщо у них співпадає хоча б один мегашингл.

Необхідно зазначити, що методи, пов'язані з обчисленням шинглів, не є ефективним при порівнянні невеликих за розміром документів.

### *Методи лексичних сигнатур*

При порівнянні документів має сенс враховувати мовну природу останніх, чого нестачає у двох наведених вище формальних алгоритмах. Урахування цієї особливості, у деяких випадках, підвищує ефективність виявлення нечітких

дублікатів або ж, за іншим визначеннями, слабо синонімічних текстів.

У цих випадках, на відмінність від алгоритму шинглів, у якості основних одиниць вимірювання використовуються слова з документів. Цей клас алгоритмів передбачає фокусування на семантичній подібності документів, не приділяючи уваги аналізу їхньої структури. Найбільш часто використовується побудова словників опорних слів і подальше порівняння словників окремих документів між собою.

У роботі В.А. Широкова<sup>62</sup> пропонується такий найпростіший критерій слабкої синонімії: якщо більше заданого відсотку ( $n$ ) тексту  $T_1$  присутні в тексті  $T_2$  в тих самих формах, якщо більше деякого відсотку ( $m$ ) пар слів, що стоять поруч в тексті  $T_1$  присутні в тексті  $T_2$  в тих самих формах, і заданий відсоток ( $k$ ) трійок слів текстів  $T_1$  і  $T_2$  збігаються, то тексти є слабо синонімічними. Числа  $m$ ,  $n$ ,  $k$  залежать від довжини досліджуваних текстів, тобто, є функціями від довжини, розміру словника текстів, жанру тощо.

Також існує декілька підходів для отримання числових значень міри близькості між документами. Якщо позначити множину опорних слів документа  $T$  як  $S(T)$ , а  $|S(T)|$  як міру появи слів в документі, що визначається як кількість опорних слів, то міри близькості обчислюються наступним чином:

$$\text{Косинус: } \text{simCos}(A, B) = \frac{|S(A) \cap S(B)|}{\sqrt{|S(A)|^2 + |S(B)|^2}};$$

---

<sup>62</sup> Широков В.А., Бугаков О.В., Грязнухіна Т.О. Корпусна Лінгвістика. – К.: Довіра, 2005. – 471 с.

$$\text{Коефіцієнт Жаккарда: } \mathit{simJaccard}(A, B) = \frac{|S(A) \cap S(B)|}{|S(A) \cup S(B)|};$$

$$\text{Коефіцієнт Дайса: } \mathit{simDais}(A, B) = \frac{|S(A) \cap S(B)|}{|S(A)| + |S(B)|}.$$

У деяких випадках ефективною є несиметрична спрощена міра близькості:

$$\mathit{simNSL}(A, B) = \frac{|S(A) \cap S(B)|}{|S(A)|},$$

а також симетрична, що обчислюється як сума двох несиметричних:

$$\mathit{simSSL}(A, B) = \frac{|S(A) \cap S(B)|}{|S(A)|} + \frac{|S(A) \cap S(B)|}{|S(B)|}.$$

У цьому випадку  $|S(A)|$  – кількість слів у тексті  $A$ ,  $S(A) \cap S(B)$  – перетин множин слів текстів  $A$  і  $B$ , тобто слова із  $A$  і  $B$ , що збігаються.

Наведений вище найпростіший критерій слабкої синонімії текстів використовує деякі числові характеристики пар текстів. Як критерій лексичної відстані двох текстів  $T_1$  і  $T_2$  у цьому випадку застосовують величину  $\mathit{simNSL}(T_1, T_2)$ . Із визначення випливає, що  $\mathit{simNSL}(T_1, T_2) \neq \mathit{simNSL}(T_2, T_1)$ ,  $\mathit{simNSL}(T_1, T_2) \leq 1$ , якщо  $T_1$  і  $T_2$  не збігаються та  $\mathit{simNSL}(T_1, T_2) = 1$ , якщо  $T_1 = T_2$ . Позначимо  $T_1^2$  – множину пар словоформ тексту  $T_1$ , що стоять поруч, а  $T_1^3$  – множину трійок словоформ, і, узагальнюючи,  $T_1^n$  – множину кортежів з  $n$  словоформ. Лексичну відстань  $n$ -го порядку визначимо таким чином:

$$simNSL_n(T_1, T_2) = \frac{|T_1^n \cap T_2^n|}{(|T_1| - (n-1))}.$$

Тоді наведений вище критерій можна переформулювати таким чином: якщо  $simNSL(T_1, T_2) < k$ ,  $simNSL_2(T_1, T_2) < m$ ,  $simNSL_3(T_1, T_2) < n$ , то тексти  $T_1$  і  $T_2$  є слабкосинонімічними.

Наведений критерій має працювати лише тоді, коли тексти мають спільні фрагменти, тобто близькі як за змістом, так і за формою. Якщо ж тексти, що порівнюються, різні за формою, пропонується застосовувати лексико-граматичну відстань між текстами  $T_1$  і  $T_2$ , що визначається формулою:

$$simLG_n(T_1, T_2) = \frac{\sum_{i=1, |T_1|, j=1, |T_2|} v_i^n v_j^n \lambda^n}{(|T_1|)},$$

де  $\lambda$  визначається в залежності від подібності слів ( $\lambda = 1$ , якщо слова співпадають  $v_i, v_j$ ,  $1/2$ , якщо слова  $v_i, v_j$  – точні синоніми тощо).

Ще один з підходів, що базується на розрахунку лексичної дактилограми, має назву I-Match і був запропонований в роботі<sup>63</sup>. Для цього для вихідної множини документів будується словник, який включає слова з середніми значеннями інверсної частоти документа IDF.

Значення IDF обчислюються за формулою:

$$IDF = \log \frac{|D|}{|t_j \in d_j|},$$

---

<sup>63</sup> Chowdhury A., Frieder O. etc. Collection statistics for fast duplicate document detection // ACM Transactions on Information Systems (TOIS), April 2002. – 20, Issue 2. – P. 171-191.

де  $|D|$  – кількість документів у колекції;  $|t_j \in d_j|$  – кількість документів, де зустрічається  $t_j$ .

Слова з середніми значеннями IDF, як правило, забезпечують більш точні результати при виявленні нечітких дублікатів. Після цього для кожного документа вибираються слова, які також входять до сформованого словника. Множина цих слів впорядковується і розраховується відповідна контрольна сума (застосовується хеш-функція SHA-1), яка й має назву I-Match. Два документи вважаються нечіткими дублікатами, якщо в них відповідні значення I-Match співпадають.

Наведений алгоритм з погляду обчислювальної складності більш ефективний, ніж алгоритм шинглів, крім того, він може застосовуватися для порівняння невеликих за розміром документів. Однак застосування жорсткої хеш-функції SHA-1 робить його нестійким до невеликих змін змісту документів.

Таким чином, порівняльний аналіз текстів може бути корисним інструментом для забезпечення консистентності і виявлення різних форм неоднозначності в законодавчих документах, що є предметом парламентського контролю.

- Одним з прикладів може бути порівняння версій законодавчих актів або проектів законів, що були внесені на розгляд до парламенту. Використовуючи інструменти аналізу текстів, можна порівняти різні версії документів та виявити відмінності між ними, наприклад, зміни в формулюваннях, додаткові або видалені положення тощо. Це може допомогти парламентарям зрозуміти, які зміни були внесені до законодавчого акту та чи вони відповідають поставленим цілям.



- Аналіз стенограм парламентських засідань. Порівняння стенограм різних засідань може допомогти виявити суперечності та розбіжності в думках парламентарів, що може бути важливим для визначення пріоритетів та напрямків роботи.
- Порівняння програм різних політичних партій. При порівнянні програм різних політичних партій можна знайти спільні та відмінні положення, що може допомогти визначити головні цілі та пріоритети політичної діяльності.
- Аналіз текстів законопроектів. Порівняння текстів законопроектів, що були внесені різними депутатами чи фракціями, може допомогти виявити різні підходи до вирішення конкретних проблем, що може бути корисним для забезпечення консистентності в законодавчому процесі.
- Порівняння законодавчих актів різних країн. При порівнянні законодавчих актів різних країн можна знайти схожості та відмінності між ними, що може допомогти визначити кращі практики та рекомендації для покращення законодавства відповідної країни.

Усі ці приклади можуть бути реалізовані з використанням різних інструментів для порівняльного аналізу текстів, таких як інформаційні системи з використанням штучного інтелекту, програмні засоби для автоматичного зіставлення текстів, тощо.

***Питання для самоконтролю:***

- Що суттєво знижує складність обчислень в методі рандомізованих дактилограм?
- Чому метод рандомізованих дактилограм дозволяє стверджувати про входження підрядку у рядок не абсолютно точно?

- Яка складність алгоритму Карпа-Рабіна і як вона залежить від довжини тексту та підрядка?
- Як можна використовувати алгоритм Карпа-Рабіна для виявлення дублікатів документів?
- Що таке шингли і як вони використовуються для оцінки подібності документів?
- Які недоліки мають алгоритми Карпа-Рабіна та шинглів у порівнянні з алгоритмами, що враховують мовну природу документів?
- Яким чином клас алгоритмів, що використовують слова як основні одиниці вимірювання, дозволяє фокусуватися на семантичній подібності документів?
- Які методи побудови словників опорних слів використовуються в цьому класі алгоритмів?
- Які є переваги та недоліки класу алгоритмів, що використовують слова як основні одиниці вимірювання?
- Як порівняльний аналіз текстів може допомогти виявити різні форми неоднозначності в законодавчих документах?
- Які є приклади використання порівняльного аналізу текстів у законодавчому процесі?
- Які можуть бути вигоди від порівняння стенограм парламентських засідань?
- Як порівняння текстів законопроектів може допомогти забезпечити консистентність в законодавчому процесі?
- Які можуть бути переваги порівняння законодавчих актів різних країн?

## **6. ШТУЧНИЙ ІНТЕЛЕКТ У ПАРЛАМЕНТСЬКОМУ КОНТРОЛІ**

### ***6.1. Поняття штучного інтелекту***

Тема штучного інтелекту останнім часом стала вкрай популярною, і маємо безліч публікацій, присвячених їй. Відзначимо надзвичайно широкий, сказати б, жанровий діапазон – від наукових праць в фахових журналах до науково-фантастичної художньої літератури. І, нажаль, далеко не завжди витримується грань, що відділяє науку від фантастики. Часто, наприклад, цілком серйозно обговорюють реальну перспективу того, що штучний інтелект зможе якщо не повністю, то значною мірою замінити людину і тим самим витіснити її з багатьох сфер суспільного життя.

Тому для правильного розуміння дійсності та реалістичної оцінки конкретних проєктів слід внести ясність в поточний стан справ щодо розвитку штучного інтелекту.

Перш за все, зазначимо, що штучний інтелект не є чимось таким, що може існувати як таке в нашому світі і мати безпосередні зв'язки з речами і явищами в ньому. Найкращим чином, на нашу думку, розуміння штучного інтелекту запропонувала компанія Oracle: це програмування комп'ютера в такий спосіб, щоб він певною мірою імітував поведінку людини. Слово «імітував» тут вимагає певних коментарів.

Штучний інтелект (англ. Artificial Intelligence, AI) – це галузь комп'ютерних наук, яка займається розробкою алгоритмів та програм, що можуть моделювати імітацію людського мислення і поведінки. Штучний інтелект вивчає розробку і застосування комп'ютерних систем, які можуть виконувати завдання, які зазвичай потребують інтелектуальної людської діяльності, такі як розпізнавання зображень, мови, прийняття рішень та ін.

Застосування штучного інтелекту в галузі права можливе і вже активно розвивається. Штучний інтелект може бути корисним для автоматизації рутинних процесів, таких як перевірка документів на відповідність законодавству, пошук та аналіз правової інформації, прогнозування рішень судів на основі аналізу судової практики, виявлення та аналіз недоліків у законодавстві, а також у підтримці прийняття рішень в реальному часі на основі аналізу великих обсягів даних.

Суб'єктом парламентського контролю в Україні є Верховна Рада України. Але виникає питання, як такий контроль може ефективно здійснюватися на практиці, враховуючи географічні розміри та суспільно-політичну складність України? Очевидно, що ключову роль тут має відіграти використання сучасних інформаційних технологій. Адже в основі парламентського контролю лежить збір і опрацювання інформації. І в цьому плані надзвичайно перспективними слід вважати системи з елементами штучного інтелекту, оскільки вони дозволяють машині не лише накопичувати і здійснювати первинну обробку «сирих» даних, але й виконувати певні операції з її обробки в тих рамках, які досі вважалися доступними лише людині. Як приклад наведемо принципову можливість на основі розвиненої обробки вхідних даних приймати певні рішення в режимі реального часу, що є вкрай важливо в процесі оперативного реагування на різноманітні ситуації, які важко помітити й адекватно проаналізувати людині.

Нижче ми обговоримо можливості використання елементів штучного інтелекту в сфері парламентського контролю під певним кутом зору, пов'язаним з агентно-орієнтованим підходом, про який ми вже згадували в попередніх розділах. Ми зосередимо увагу на тій характерній особливості концепції інтелектуального агента, яка полягає в тому, що він

знаходиться в стані інформаційного обміну з своїм зовнішнім середовищем. Нас цікавитиме структура і властивості середовища, в якому перебуватимуть інтелектуальні агенти, задіяні в парламентському контролі. Нагадаємо, що на загальному рівні під зовнішнім середовищем інтелектуального агента розуміють все, що може бути джерелом вхідних даних для нього. Але для глибшого розуміння поняття «зовнішнє середовище» слід уточнити.

Будь-яка комп'ютерна система в певному розумінні знаходиться в стані інформаційного обміну з оточуючим середовищем. Отже, якщо ми говоримо про штучний інтелект, необхідно додати ряд умов до розуміння як середовища, так і обміну інформацією між ним і машиною, тим самим звизивши відповідні поняття.

Часто кажуть, що специфіка дії інтелектуального агента полягає в тому, що він «розуміє» стан, в якому знаходиться середовище і він сам, а тому його дії мають певну мету, на досягнення якої спрямовані його дії. Проблемою, однак, є те, що досі відсутнє строге визначення поняття «розуміти», внаслідок чого неможливо однозначно визначити, розуміє машина щось чи ні. Не набагато краще в цьому плані виглядає і поняття «мета». Скажімо, чи можна вважати, що машина, обчислюючи факторіал, розуміє, що вона робить? І чи можна вважати, що метою цього процесу є визначення його значення? В усякому разі, якби мова йшла про людину, обидві відповіді, очевидно, були б позитивними. Адже людина не може обчислювати факторіал, якщо не розуміє що це таке і як робиться. Але обчислити факторіал може найпримітивніший комп'ютер, щодо якого питання про штучний інтелект взагалі не стоїть.

Значно складнішою є, наприклад така ситуація. Машина отримує вхідні дані, в структурі яких набір параметрів виходить за межі припустимих значень, а в такому випадку

система повинна виконати певні дії. Вони можуть бути здійснені різними засобами, при чому машина сама обирає один із них, керуючись заданими критеріями. В процесі виконання може бути реалізований програмний блок, що належить до набору процедур, передбачених загальними алгоритмами. І лише частина цих процедур має сенс з точки зору виконання машиною необхідних в даній ситуації дій. В тому числі це може бути обчислення факторіалу, який входить як множник до складнішої формули. Тоді, оскільки машина обчислює факторіал за власним вибором тому, що в такий спосіб може бути отриманий бажаний результат, можемо говорити про певне розуміння ситуації машиною. Звичайно, наведений приклад надто примітивний і схематичний, але він ілюструє характер центральної проблеми, що виникає у зв'язку з поняттям «розуміння».

Щоб зробити наступний крок, зазначимо, що коли йдеться про інтелектуальних агентів, то під «інтелектуальністю» в першу чергу мається на увазі наявність у машини зворотних зв'язків між різними програмними компонентами заснованих на результатах аналізу різноманітних даних, як вхідних, так і проміжних, що отримуються в процесі роботи апаратно-програмного комплексу. Для нас важливим є перш за все те, що принаймні деякі зворотні зв'язки працюють відповідно до даних, отриманих від оточення системи. Іншими словами, в таких випадках машина не виконує однозначні детерміновані дії, передбачені закладеними в неї алгоритмами, а приймає рішення, що ґрунтується на її внутрішній роботі. Саме так на загальному рівні працює людський розум. Досі не відомі його механізми, але є підстави вважати, що вони якісно відмінні від механізмів в комп'ютері, але з операціоналістичної точки зору мають структурну подібність.

Штучний інтелект може вирішувати багато задач в правовій інформатиці, зокрема:

1. Класифікація інформації: ШІ може автоматично класифікувати документи за темою, типом, структурою і т.д. Це забезпечує більш швидкий і точний пошук інформації.
2. Пошук інформації: ШІ може використовувати різні алгоритми пошуку інформації, такі як розумний пошук, обробка природної мови і т.д. Це дозволяє швидко знаходити необхідну інформацію в масиві даних.
3. Автоматичне створення документів: ШІ може використовувати тексти з різних джерел для створення нових документів. Наприклад, він може створити проект закону на основі аналізу різних документів і матеріалів.
4. Аналіз даних: ШІ може використовувати алгоритми машинного навчання для аналізу великих масивів даних. Наприклад, він може використовувати аналіз даних, щоб знайти кореляції між різними законами і політиками.
5. Розпізнавання образів і голосів: ШІ може використовувати технології розпізнавання образів і голосів, щоб ідентифікувати людей і автоматизувати процес реєстрації відвідувачів інформаційних ресурсів.

Таким чином, штучний інтелект може вирішувати складні завдання в правовій інформатиці, що дозволяє покращити якість та швидкість, зокрема, парламентського контролю.

Протягом досить довгого часу вважалося самоочевидним, що штучний інтелект повинен мати здатність мислити в тому ж сенсі, що й людина. В спрощеному формулюванні це означає, що штучний інтелект – це машина, що здатна мислити. Квінтесенцією такого погляду став тест Тюрінга, який на думку автора (і багатьох дослідників) дозволяв

визначити, чи дійсно дана машина є штучним інтелектом. Цей тест дуже простий: відповідь вважається позитивною, якщо людина в процесі спілкування не може визначити, з ким вона спілкується – з машиною чи іншою людиною.

На перший погляд тест виглядає цілком логічним. Але швидко виявилися проблеми, в тому числі концептуального рівня. Зупинимося на одній з них, яка отримала назву «Універсальний грамофон». Уявімо собі базу даних, яка містить всі можливі запитання, які здатна поставити людина. Кожному з них відповідає набір осмислених відповідей. Тоді людина, що приймає участь в тесті Тюринга, на кожне своє запитання отримує відповідь, яку могла дати інша людина. Очевидно, що робота такої машини не має нічого спільного з людським мисленням, але при цьому демонструє людську поведінку. Звичайно, «Універсальний грамофон» теж вразливий для критики, але ми не будемо заглиблюватися в глибокий аналіз даної ситуації. Для нас важливо, що за певних умов можемо мати імітацію поведінки людини, що є лише формальною (зовнішньою) і тому придатною лише для демонстрації окремих технічних рішень.

Спроба розширити методику тестування також не знімає заперечень. Так, наприклад, на перший погляд можна було б включити в тест окрім гри у відповіді на запитання також розв'язання задач. Але на принциповому рівні це нічого не змінює вже хоча б тому, що завжди знайдеться людина, яка не зможе розв'язати дану задачу. А певний клас задач може бути включений до нашої бази даних, отже машина знайде (причому в буквальному розумінні) правильний розв'язок.

Та й взагалі не варто забувати, що звичайна людина часто робить досить сумнівні висловлювання.

Ключовою проблемою є те, що ми досі дуже мало знаємо про людське мислення та людську поведінку. Тому, якщо



штучний інтелект розуміти як результат моделювання процесів мислення людини, то не зрозуміло, що власне треба моделювати. Також заздалегідь невідомо, чи це взагалі потрібно в плані розвитку інформаційних технологій. Навіть отримавши відповідь (позитивну чи негативну) на філософське питання «Чи здатна машина мислити?», ми можемо не отримати реального поступу в розвитку комп'ютерних систем. Також відкритим лишається питання, чи потрібен нам штучний інтелект, який нічим не відрізняється від природного.

Поступово почала домінувати думка, що за своєю структурою та характером діяльності взагалі не повинен відтворювати природний. Мета повинна полягати не в тому, щоб штучними засобами створити інтелект, подібний людському, а в тому, щоб отримати можливість без участі людини вирішувати задачі, які не вирішуються «звичайними» комп'ютерами. Як машина буде це робити, не має значення. Важливо зазначити, що штучний інтелект концептуально не здатен виконувати ті чи інші операції самостійно, «з власної ініціативи». Можливість їх виконання має бути в той чи інший спосіб передбачена програмою, яку створює людина. Інша справа, що такі дії машини можуть не бути однозначно детерміновані загальним алгоритмом. З точки зору зовнішнього спостереження це може виглядати як самостійне прийняття рішення машиною, і до певної міри так воно і є. Але насправді машина просто вибирає один з передбачених варіантів досягнення заданого результату. За певних умов машина може навіть модифікувати і вдосконалити наявний варіант відповідно до закладеної в неї моделі. Та «вгадати» якісно новий варіант, не передбачений розробниками, машина принципово не здатна.

З іншого боку, слід розуміти, що така поведінка може зробити діяльність штучного інтелекту непередбачуваною, і

це створює специфічні проблеми, які ми зараз не будемо обговорювати. В будь-якому випадку машина не може вийти за межі скінченого простору можливих операцій, і в цьому полягає фундаментальна відмінність штучного інтелекту від природного, який має творчий початок.

Окреме місце займає питання вхідних даних для систем штучного інтелекту. Людина отримує їх через перцептивну систему (органи відчуття). Отже, якщо ставиться задача моделювання можливостей людини, штучний інтелект повинен мати якийсь аналог такої системи. Тобто штучний інтелект повинен приймати інформацію, надану відеокамерами, мікрофонами тощо і обробляти її подібно до того, як це робить людина. Але якщо від штучного інтелекту вимагається лише отримати розв'язок заданої задачі, природа і структура вхідних даних не має принципового значення. В принципі, штучний інтелект може отримувати вхідні дані безпосередньо в цифровому форматі, підготовленому до подальшої обробки. Таким чином, частина операцій (з підготовки вхідних даних) може виконуватись окремо, в тому числі за безпосередньою участю людини. З прагматичної точки зору такий підхід виглядає значно перспективнішим.

Створення комп'ютерних систем з елементами штучного інтелекту для рішення задач правової інформатики має великий потенціал для покращення ефективності та точності обробки правової інформації. Застосування штучного інтелекту в правовій сфері ще не досягло свого потенціалу, проте деякі прогресивні технології, такі як машинне навчання, глибоке навчання та нейронні мережі, вже успішно використовуються для вирішення певних задач.

Інтелектуальні агенти можуть бути використані для автоматизації виконання повторюваних процесів, таких як робота з текстовими документами та аналіз великих обсягів даних. Вони також можуть бути використані для побудови

систем експертів, які допомагають користувачам з рішенням складних правових питань.

Перспективним напрямком є розробка систем, які можуть автоматично аналізувати та інтерпретувати правові документи, зокрема законодавчі акти та рішення суду, та надавати рекомендації користувачам щодо їх застосування.

Також перспективною є розробка систем, які можуть працювати з мовою людей та здійснювати пошук та аналіз інформації з метою вирішення правових питань.

Однією з перспектив є створення інтелектуальних агентів, які зможуть здійснювати розумний пошук та аналіз правової інформації, підбирати необхідні матеріали для рішення конкретної задачі, давати рекомендації та прогнози, а також допомагати у роботі з правовими документами.

Крім того, можливість автоматизації процесу обробки правової інформації за допомогою штучного інтелекту може допомогти зменшити кількість помилок та прискорити процес прийняття рішень. Наприклад, системи розпізнавання тексту та аналізу мови можуть допомогти автоматично класифікувати документи за їх тематикою та рівнем важливості.

Крім того, штучний інтелект може бути використаний для створення систем прогнозування рішень, які будуть базуватися на аналізі великих обсягів правової інформації та статистичних даних. Це допоможе зменшити ризики та підвищити ефективність роботи в галузі права.

Отже, створення комп'ютерних систем з елементами штучного інтелекту для рішення задач правової інформатики має великий потенціал і може внести значний внесок у розвиток цієї галузі.

### ***Питання для самоконтролю:***

- Які завдання можуть бути автоматизовані за допомогою штучного інтелекту в галузі права?
- Які переваги можуть бути отримані завдяки застосуванню штучного інтелекту в галузі права?
- Яким чином можуть комп'ютерні системи на основі штучного інтелекту аналізувати судову практику та прогнозувати рішення судів?
- Як штучний інтелект може виявляти недоліки у законодавстві?
- Чому використання штучного інтелекту в галузі права є актуальним і важливим напрямом розвитку?
- Які технології штучного інтелекту успішно використовуються для вирішення певних задач у правовій сфері?
- Які задачі можуть бути автоматизовані за допомогою інтелектуальних агентів в правовій сфері?
- Які можливості надає розробка систем, які автоматично аналізують та інтерпретують правові документи?
- Які переваги можуть бути забезпечені автоматизацією процесу обробки правової інформації за допомогою штучного інтелекту?
- Як можуть бути використані системи прогнозування рішень на основі аналізу великих обсягів правової інформації та статистичних даних у правовій сфері?

### ***6.2. Інтелектуальні агенти***

Коли ми говоримо, що штучний інтелект імітує діяльність людини, то це слід розуміти так, що він здатен розв'язувати задачі, які традиційно відносилися до компетенції людини.

При цьому може виникати свого роду ієрархія задач – машині доводиться розв'язувати задачі різних рівнів по відношенню одна до одної. Верхній рівень становить задача, яку перед машиною ставить людина. Але для її розв'язку може знадобитись розв'язати деякий набір задач другого рівня. Кожна задача другого рівня може передбачати розв'язок задач третього рівня і т. д. коли йдеться про штучний інтелект, допускається можливість (принаймні гіпотетично), що частину таких задач нижчих рівнів ставитиме сама машина без участі людини. Звичайно, робитиме вона це відповідно до наборів моделей і правил, заданих людиною. Але ця обставина може зробити діяльність машини надзвичайно складною. І цим штучний інтелект якісно відрізняється від звичайного комп'ютера, яким би складним і досконалим він не був.

Такі можливості зумовили появу агентно-орієнтованого підходу до створення штучного інтелекту, в якій ключову роль відіграє концепція Інтелектуального агента в довкіллі. Останнім часом цей підхід набуває все більшу популярність, оскільки органічно поєднує цілий ряд окремих рис, властивих іншим підходам.

Ще з 90-х років ХХ століття інтенсивно розвивається напрямок вивчення проблеми штучного інтелекту, заснований на агентно-орієнтованому підході. Центральним для нього є поняття інтелектуального агента.

Під інтелектуальним агентом розуміють кібернетичну сутність, яка знаходиться в стані інформаційного обміну з своїм оточенням, приймає рішення відповідно до отриманої інформації і діє для досягнення конкретної цілі. Оточення в даному випадку - це все, що може бути джерелом вхідних даних для інформаційного агента.

З технологічної точки зору інтелектуальний агент є програмою (або комп'ютерною системою, що має таку програму), яка працює постійно, не вимагаючи від персоналу відповідно до потреб вмикання й вимикання. Агент автоматично отримує вхідні дані по передбаченим каналам (в тому числі від сенсорних систем), здійснює їх обробку і виконує належні дії. При цьому він виконує, якщо це потрібно, керування певними процесами. Тому, на наш погляд, використання інтелектуальних агентів є перспективним в різноманітних процесах контролю, в тому числі і парламентського.

Однією з характерних особливостей концепції інтелектуального агента є те, що він знаходиться в стані інформаційного обміну з своїм зовнішнім середовищем. Нижче ми зосередимось саме на цьому аспекті агентно-орієнтованого підходу в рамках парламентського контролю. Сфера здійснення парламентського контролю може розглядатися як певне середовище, інформаційний обмін з яким може забезпечити ефективну роботу інтелектуальних агентів. А це, в свою чергу, визначає один з можливих напрямків використання в даній сфері штучного інтелекту. Особливу увагу ми приділимо ролі зворотних зв'язків в інформаційному середовищі.

Концепція інтелектуального агента базується на ідеї створення програмного забезпечення, яке може взаємодіяти з оточуючим середовищем та виконувати завдання з мінімальною людською участю. Інтелектуальний агент - це програмний засіб, який може сприймати, аналізувати і реагувати на певні події у своєму середовищі, збирати та обробляти інформацію, здійснювати прийняття рішень і виконувати дії відповідно до поставленої мети.

Інтелектуальні агенти можуть застосовуватись в різних галузях, включаючи правову інформатику. В контексті

правової інформатики, інтелектуальні агенти можуть допомогти в зборі та аналізі юридичної інформації, автоматизації процесу прийняття рішень, підтримці досліджень у галузі права та багато іншого. Використання інтелектуальних агентів може допомогти покращити ефективність та точність роботи в галузі правової інформатики, зменшити час і витрати на обробку інформації та забезпечити більш швидкий та точний доступ до юридичної інформації.

Під інтелектуальним агентом в галузі штучного інтелекту розуміють комп'ютерну систему, яка отримує інформацію про стан певних процесів через набір сенсорів і здійснює на них вплив, який полягає в приведенні деякого набору параметрів до заданих значень. З сучасної точки зору, поширеної серед фахівців в галузі штучного інтелекту, термін "інтелектуальний" щодо агента слід розуміти не як твердження про наявність у нього інтелекту в звичайному розумінні, а як вищий рівень технології в порівнянні з простими тригерними системами. Основним критерієм визначення системи як інтелектуального агента є адекватність його взаємодії з оточенням тому чи іншому набору вимог. Часто також в цьому сенсі вживають вираз "раціональність" поведінки. Дослідження, пов'язані з інтелектуальними агентами в основному ведуться в різноманітних напрямках на стику інформатики та інших галузей науки і технології.

Для нас також важливими є дослідження середовища, в якому діє агент.

Взагалі, середовище визначається як сукупність умов існування людини та суспільства. У соціології під соціальним середовищем людини розуміють економічні, політичні, соціальні, духовні, територіальні умови, що впливають на становлення особистості. Нас цікавитиме інформаційне

середовище комп'ютерної системи яке може бути визначене як сукупність умов, в якій вона функціонує, а також стан її локальних підсистем та глобальних систем, по відношенню до яких вона сама є підсистемою. В галузі вивчення його структури та властивостей провадилися численні дослідження, і даний напрямок досліджень є досить глибоко і детально вивчений.

Структура інформаційного середовища на загальному рівні виглядає наступним чином:

- суб'єкти інформаційної взаємодії;
- інформація, призначена для використання суб'єктами інформаційної взаємодії
- інформаційна інфраструктура, що забезпечує обмін інформацією між суб'єктами інформаційної взаємодії;
- суспільні відносини, які виникають в процесі генерації, зберігання, модифікації та використання інформації.

Саме інформаційне середовище може розглядатись як основа середовища, в якому функціонують інтелектуальні агенти.

Таким чином, під інтелектуальним агентом розуміють кібернетичну сутність, яка знаходиться в стані інформаційного обміну з своїм оточенням, приймає рішення відповідно до отриманої інформації і діє для досягнення конкретної цілі. Оточення в даному випадку – це все, що може бути джерелом вхідних даних для інформаційного агента.

І в цьому відношенні інтелектуальний агент дійсно до певної міри імітує поведінку людини.

Розрізняють два види інтелектуальних агентів:

- фізичні;



- часові.

Фізичні агенти отримують вхідні дані від сенсорних систем і діють за допомогою маніпуляторів, тим самим безпосередньо впливаючи на реальні процеси. Часові агенти отримують інформацію і генерують пропозиції щодо конкретних дій інших машин чи людини.

За способом обробки вхідних даних інтелектуальні агенти розподіляються на такі категорії:

- агенти з простою поведінкою;
- агенти з поведінкою, заснованою на моделі;
- цілеспрямовані агенти;
- практичні агенти;
- агенти, що навчаються.

#### ***Агенти з простою поведінкою***

Агенти з простою поведінкою діють тільки на основі поточних знань. Їхня агентська функція заснована на схемі умова-дія

IF (умова) THEN дія

Така функція може бути успішною, тільки якщо навколишнє середовище повністю піддається спостереженню. Деякі агенти також можуть мати інформацію про їхній поточний стан, що дозволяє їм не звертати уваги на умови, передумови яких уже виконані.

#### ***Агенти з поведінкою, заснованою на моделі***

Агенти з поведінкою, заснованою на моделі, можуть оперувати із середовищем, що лише частково піддається спостереженню. Усередині агента зберігається уявлення про ту частину, що перебуває поза межами огляду. Щоб мати таке уявлення, агентові необхідно знати, як виглядає навколишній

світ, як він улаштований. Ця додаткова інформація доповнює «Картину Світу».

### ***Цілеспрямовані агенти***

Цілеспрямовані агенти схожі з попереднім типом, однак вони, крім іншого, зберігають інформацію про ті ситуації, які для них бажані. Це дає агентові спосіб вибрати серед багатьох шляхів той, що приведе до потрібної мети.

### ***Практичні агенти***

Цілеспрямовані агенти розрізняють тільки стани, коли ціль досягнута, і коли не досягнута. Практичні агенти, крім цього, здатні розрізняти, наскільки бажаний для них поточний стан. Така оцінка може бути отримана за допомогою «функції корисності», що проектує множину станів на множину мір корисності станів.

### ***Агенти, що навчаються***

Агенти, що навчаються (АН), також називаються автономними інтелектуальними агентами (англ. *autonomous intelligent agents*), що означає їхню незалежність і здатність до навчання й пристосовування до обставин, що змінюються. Система АН повинна виявляти такі здатності:

- навчатися й розвиватися в процесі взаємодії з навколишнім; середовищем;
- пристосовуватися в режимі реального часу;
- швидко навчатися на основі великого обсягу даних;
- покроково пристосовувати нові способи розв'язання проблем;
- володіти базою прикладів із можливістю її поповнення;
- мати параметри для моделювання швидкої й довгої пам'яті, віку й т.д.
- аналізувати себе у термінах поведінки, помилки й успіху.

Слід зауважити, що система штучного інтелекту, взагалі кажучи, може суміщати в собі елементи різних категорій.

Ми зосередимо увагу на тій характерній особливості концепції інтелектуального агента, яка полягає в тому, що він знаходиться в стані інформаційного обміну з своїм зовнішнім середовищем. Нас цікавитиме структура і властивості середовища, в якому перебуватимуть інтелектуальні агенти, задіяні в парламентському контролі. Нагадаємо, що на загальному рівні під зовнішнім середовищем інтелектуального агента розуміють все, що може бути джерелом вхідних даних для нього. Але для глибшого розуміння поняття "зовнішнє середовище" слід уточнити.

Будь-яка комп'ютерна система в певному розумінні знаходиться в стані інформаційного обміну з оточуючим середовищем. Отже, якщо ми говоримо про штучний інтелект, необхідно додати ряд умов до розуміння як середовища, так і обміну інформацією між ним і машиною, тим самим звузивши відповідні поняття.

Часто кажуть, що специфіка дії інтелектуального агента полягає в тому, що він «розуміє» стан, в якому знаходиться середовище і він сам, а тому його дії мають певну мету, на досягнення якої спрямовані його дії. Проблемою, однак, є те, що досі відсутнє строге визначення поняття «розуміти», внаслідок чого неможливо однозначно визначити, розуміє машина щось чи ні. Не набагато краще в цьому плані виглядає і поняття «мета». Скажімо, чи можна вважати, що машина, обчислюючи факторіал, розуміє, що вона робить? І чи можна вважати, що метою цього процесу є визначення його значення? В усякому разі, якби мова йшла про людину, обидві відповіді, очевидно, були б позитивними. Адже людина не може обчислювати факторіал, якщо не розуміє що це таке і як робиться. Але обчислити факторіал може

найпримітивніший комп'ютер, щодо якого питання про штучний інтелект взагалі не стоїть.

Значно складнішою є, наприклад така ситуація. Машина отримує вхідні дані, в структурі яких набір параметрів виходить за межі припустимих значень, а в такому випадку система повинна виконати певні дії. Вони можуть бути здійснені різними засобами, при чому машина сама обирає один із них, керуючись заданими критеріями. В процесі виконання може бути реалізований програмний блок, що належить до набору процедур, передбачених загальними алгоритмами. І лише частина цих процедур має сенс з точки зору виконання машиною необхідних в даній ситуації дій. В тому числі це може бути обчислення факторіалу, який входить як множник до складнішої формули. Тоді, оскільки машина обчислює факторіал за власним вибором тому, що в такий спосіб може бути отриманий бажаний результат, можемо говорити про певне розуміння ситуації машиною. Звичайно, наведений приклад надто примітивний і схематичний, але він ілюструє характер центральної проблеми, що виникає у зв'язку з поняттям "розуміння".

Щоб зробити наступний крок, зазначимо, що коли йдеться про інтелектуальних агентів, то під "інтелектуальністю" в першу чергу мається на увазі наявність у машини зворотних зв'язків між різними програмними компонентами заснованих на результатах аналізу різноманітних даних, як вхідних, так і проміжних, що отримуються в процесі роботи апаратно-програмного комплексу.

В рамках аналізу штучного інтелекту одним із ключових є поняття нелінійної системи, що, в свою чергу означає, що вона здатна виконувати нетривіальні дії, які принципово не зводяться до простого виконання алгоритмів, закладених в машину. І ця обставина створює умови, необхідні для того, щоб система могла бути застосована в сфері парламентського

контролю. Важливо, що в даному випадку нелінійність системи головним чином зумовлена наявністю в ній зворотних зв'язків. Саме вони дають підстави говорити про можливість використання машини з метод здійснення контрольних функцій.

Зворотні зв'язки є основою кібернетики та одним із ключових понять теорії систем. На загальному рівні зворотнім зв'язком називають вплив результатів деякого керованого процесу на його протікання. Повна теорія зворотних зв'язків досить складна і далеко виходить за межі нашої теми. Тому ми не будемо торкатися математичних питань, обмежившись лише загальними уявленнями.

В найпростішій реалізації зворотній зв'язок полягає в тому, що на вхід деякого пристрою подається сигнал, який є функцією його вихідного сигналу. Зворотні зв'язки можуть застосовуватись з різними цілями, але головне їх призначення полягає в забезпеченні можливості змінювати режим роботи пристрою залежно від змін вихідного сигналу. Відповідно існує два види зворотних зв'язків: додатні і від'ємні. Додатні зворотні зв'язки посилюють зміну вихідного сигналу, а від'ємні - змінюють вихідний сигнал так, щоб протидіяти його зміні.

Від'ємні зворотні зв'язки використовуються як стабілізуючі елементи системи, які забезпечують (наскільки це можливо) стабільність її роботи. Додатні зворотні зв'язки здебільшого використовують як генератори.

Але зворотні зв'язки існують не лише в приладах, створених людиною. Вони виникають стихійно в будь-яких складних системах, в тому числі суспільних. І повна картина певного процесу повинна, взагалі кажучи, враховувати всі зворотні зв'язки.

Роль зворотних зв'язків в системах штучного інтелекту розглядалась нами раніше. Зокрема зазначалось, що принаймні деякі зворотні зв'язки працюють відповідно до даних, отриманих від оточення системи. Іншими словами, в таких випадках машина не виконує однозначні детерміновані дії, передбачені закладеними в неї алгоритмами, а приймає рішення, що ґрунтується на її внутрішній роботі. Саме так на загальному рівні працює людський розум. Нам не відомі його механізми, і ми маємо підстави вважати, що вони якісно відмінні від механізмів в комп'ютері, але з операціоналістичної точки зору маємо структуру подібність.

Це дає нам підстави стверджувати, що принаймні одна принципова відмінність між взаємодією з оточенням штучного інтелекту та «звичайних» комп'ютерних систем полягає в активному використанні зворотних зв'язків. Розглядаємо зворотні зв'язки в ширшому контексті, включаючи оточення комп'ютерної системи, з яким вона знаходиться в стані постійного обміну інформацією. А саме, розглядаючи середовище інтелектуального агента, ми зосередимося саме на зворотних зв'язках. В першу чергу нас цікавитиме таке середовище в правовій сфері.

Якщо розглядати інформаційне середовище як цілісну систему, то середовище, що цікавить нас, можна визначити як її підсистему, що має набір сталих характеристик, пов'язаних з особливостями нормативно-правової інформації. З іншого боку, оскільки інтелектуальний агент існує та функціонує в суспільстві, його властивості його середовища значною мірою визначаються відповідною соціальною структурою. Нагадаємо, що соціальна структура - це внутрішній устрій суспільства або соціальної групи, що складається з певним чином розташованих, впорядкованих частин (соціальних індивідів, осіб, спільнот тощо), що знаходяться у соціальному зв'язку, а отже взаємодіють між собою. Іншими словами, це

взаєморозміщення та певний взаємозв'язок складових частин суспільства, як цілісної системи.

Тому можемо сказати, що середовище, в якому діє інтелектуальний агент, є інформаційним середовищем, асоційованим з певною соціальною структурою.

В основі зворотних зв'язків, які становлять для нас інтерес, в першу чергу лежить взаємний вплив суспільних процесів та нормативно-правової інформації. Дійсно, зміни в соціальній поведінці членів суспільства викликають необхідність змін правового регулювання, що в свою чергу зумовлює зміни нормативно правової інформації. Оновлена нормативно-правова інформація викликає відповідні зміни в соціальній поведінці, і це відбувається постійно. Ці зміни в нашому випадку відстежує інтелектуальний агент і, залежно від характеру ситуації, виконує ті чи інші дії. Типовими діями, очевидно, мають бути повідомлення про виявлені відхилення від прогнозованої норми відповідним організаціям, що виконують парламентський контроль. Оскільки подібний моніторинг правової ситуації вимагає отримання та опрацювання величезних обсягів інформації, виконувати його "вручну" є практично нездійсненою задачею. Використовуючи виключно людський ресурс, ми будемо змушені обмежитися лише порівняно незначною кількістю найбільш виразних і очевидних проблем. Проста автоматизація може принести незначне підвищення ефективності, оскільки за кожним виявленим відхиленням потрібно буде приймати рішення конкретній людині. Тоді як штучний інтелект, здатний опрацьовувати практично всю наявну інформацію і самостійно здійснювати її аналіз (який в ряді випадків завершується прийняттям рішення) здатен перевести процес контролю на якісно вищий рівень.

Таким чином, ми бачимо, що парламентський контроль під певним кутом зору може бути описаний як комплекс процесів

в специфічному середовищі. Це середовище по суті є результатом синтезу соціального середовища та інформаційного середовища і характеризується наявністю зворотних зв'язків, що надають йому цілісність. Власне парламентський контроль з цієї точки зору передбачає стабільне підтримання рівноваги між компонентами описаного вище середовища. Така рівновага може бути забезпечена шляхом використання комп'ютерних систем з елементами штучного інтелекту в рамках агентно-орієнтованого підходу. А саме, інтелектуальні агенти, взаємодіючи з цим середовищем, реалізують зворотні зв'язки, які і забезпечують належну рівновагу.

Такі технології, на наш погляд, здатні суттєво підвищити рівень автоматизації парламентського контролю.

Штучний інтелект може відігравати важливу роль в парламентському контролі, допомагаючи збирати, аналізувати та інтерпретувати великі обсяги даних, що збираються під час проведення різноманітних перевірок та розслідувань. Зокрема, за допомогою штучного інтелекту можна швидко і точно аналізувати велику кількість документів, зокрема законопроектів, звітів, доповідей, кореспонденції та інших матеріалів.

Штучний інтелект також може допомогти забезпечити точність та об'єктивність даних, що збираються під час контролю, шляхом використання алгоритмів машинного навчання та інших технологій. Крім того, застосування штучного інтелекту може допомогти забезпечити більш ефективне використання ресурсів та скоротити час, необхідний для проведення різноманітних дій у процесі парламентського контролю.

Неформалізовані задачі обробки правової інформації полягають у вирішенні проблем, які не можна вирішити за



допомогою формальних або стандартизованих методів. Такі задачі зазвичай пов'язані з відсутністю або недостатністю правової інформації, неоднозначністю її тлумачення або потребою в експертній оцінці.

Наприклад, неформалізовані задачі можуть виникати при вирішенні складних юридичних питань, які потребують врахування великої кількості факторів, в тому числі історичних, культурних та соціальних аспектів. Також неформалізовані задачі можуть виникати при забезпеченні безпеки даних, захисті інтелектуальної власності та прав людини в онлайн-середовищі.

Розв'язання неформалізованих задач у правовій інформатиці часто вимагає використання експертних знань та інтуїції, а також врахування етичних та соціальних аспектів.

Штучний інтелект може відіграти важливу роль у вирішенні неформалізованих задач обробки правової інформації. Оскільки такі задачі зазвичай пов'язані з розумінням природної мови, аналізом контексту, визначенням зв'язків та взаємозв'язків між різними елементами правової інформації, то штучний інтелект здатний допомогти в цьому процесі за рахунок використання методів машинного навчання, обробки природної мови та інших технологій.

Наприклад, штучний інтелект може допомогти у вирішенні таких задач, як автоматичне індексування текстів правових документів, виявлення ключових термінів та понять, розпізнавання іменованих сутностей (наприклад, назв організацій або осіб), аналіз контексту та зв'язків між різними елементами правової інформації, створення рекомендаційних систем для вибору правових документів тощо.

Таким чином, штучний інтелект може допомогти значно збільшити ефективність та точність обробки правової

інформації, що є особливо важливим у умовах використання великих обсягів даних в складних правових процесах.

Зокрема, системи інтелектуальних агентів можуть відігравати важливу роль в вирішенні неформалізованих задач обробки правової інформації. Завдяки своїм здібностям до аналізу та інтерпретації текстів, агенти можуть бути використані для автоматичної обробки правової інформації, що міститься у великих обсягах документів.

Системи інтелектуальних агентів можуть використовуватись для аналізу змісту документів, розпізнавання та класифікації юридичної інформації, автоматичного формування звітів та аналітичних матеріалів. Вони можуть також допомагати в розробці рекомендацій та прийнятті рішень на основі аналізу правових документів та даних. Системи інтелектуальних агентів також можуть бути корисними для автоматизації процесу збору, обробки та аналізу правової інформації, що надходить з різних джерел. Наприклад, агенти можуть бути налаштовані для автоматичного збору даних з веб-сайтів, баз даних та інших джерел, а також для перетворення цих даних у структуровану форму для подальшої обробки. Отже, системи інтелектуальних агентів можуть допомогти вирішувати неформалізовані задачі обробки правової інформації та забезпечити більш ефективну та точну обробку великих обсягів документів та даних.

Таким чином, ми бачимо, що парламентський контроль під певним кутом зору може бути описаний як комплекс процесів в специфічному середовищі. Це середовище по суті є результатом синтезу соціального середовища та інформаційного середовища і характеризується наявністю зворотних зв'язків, що надають йому цілісність. Власне парламентський контроль з цієї точки зору передбачає стабільне підтримання рівноваги між компонентами

описаного вище середовища. Така рівновага може бути забезпечена шляхом використання комп'ютерних систем з елементами штучного інтелекту в рамках агентно-орієнтованого підходу. А саме, інтелектуальні агенти, взаємодіючи з цим середовищем, реалізують зворотні зв'язки, які і забезпечують належну рівновагу.

Такі технології здатні суттєво підвищити рівень автоматизації парламентського контролю.

***Питання для самоконтролю:***

- Що таке інтелектуальний агент і які функції він виконує?
- Як інтелектуальний агент взаємодіє з оточуючим середовищем?
- Які переваги має використання інтелектуальних агентів в процесах контролю, зокрема в парламентському контролі?
- Які можливі напрямки використання штучного інтелекту в сфері парламентського контролю?
- Яка роль зворотних зв'язків в інформаційному середовищі для інтелектуальних агентів?
- Які є види інтелектуальних агентів?
- Яка відмінність між фізичними та часовими агентами?
- Як інтелектуальні агенти обробляють вхідні дані?
- Які категорії інтелектуальних агентів існують за способом їх поведінки?
- Як агенти з поведінкою, заснованою на моделі, можуть оперувати з середовищем, що лише частково піддається спостереженню?
- Яким чином у агента зберігається уявлення про ту частину світу, що перебуває поза межами огляду?

- Які основні відмінності між цілеспрямованими та практичними агентами?
- Як функція корисності допомагає практичним агентам оцінювати поточний стан?
- Які здатності повинна виявляти система агентів, що навчаються?
- Що таке нелінійна система в контексті аналізу штучного інтелекту?
- Які можливості надає наявність зворотних зв'язків в нелінійній системі для її застосування в сфері парламентського контролю?
- Яким чином працює зворотній зв'язок в найпростішій реалізації?
- Яке призначення зворотних зв'язків в системах?
- Які види зворотних зв'язків існують та в яких випадках вони застосовуються?
- Які можуть бути наслідки обмеження людського ресурсу для парламентського контролю?
- Яким чином штучний інтелект може покращити процес парламентського контролю?
- Як впливає середовище на парламентський контроль?
- Які є ключові компоненти парламентського контролю?
- Які технології можуть підвищити рівень автоматизації парламентського контролю?

### ***6.3. Проблема контролю машинної обробки даних в праві***

Варто зазначити, що штучний інтелект не може повністю замінити людський фактор у процесі прийняття рішень та

аналізу даних. Використання штучного інтелекту в парламентському контролі потребує належного навчання та компетентності фахівців, що займаються цими питаннями.

Вище ми розглядали загальні питання, пов'язані з неформалізованими задачами в праві. Тепер ми обговоримо неформалізовані задачі, які породжуються специфічними негативними чинниками, що відіграють в праві, нажаль, суттєву роль. Зупинимось на двох таких чинниках:

- суб'єктивізм;
- зловживання.

Підкреслимо, що суб'єктивізм, на нашу думку, також слід вважати негативним чинником, оскільки він може призводити до спотворення картини тієї чи іншої ситуації, а від цього залежать наслідки дій осіб та організацій, від чого часто залежать людські долі. І те, що особа, яка припустилася суб'єктивізму, не мала лихих намірів і діяла (як їй здавалося) виключно в найкращих цілях, по суті справи нічого не змінює. Але людині властивий суб'єктивізм, і боротися з ним марно. Тому ми повинні враховувати, наскільки це можливо, принципову наявність суб'єктивізму використовувати механізми, спрямовані на його компенсацію.

Такі механізми тією чи іншою мірою здійснюють порівняння різних описів даної ситуації, створених різними джерелами. Кожен з них неминуче містить елементи суб'єктивізму. Але в різних описах вони різні. Тому є підстави вважати, що в розумному наближенні перетин семантичних структур цих описів буде відповідати реальності. І чим більше таких описів ми порівняємо, тим точніший отримаємо результат. В цьому плані, очевидно, саме Big Data відкриває широкі перспективи. Ефективне використання великих обсягів різномірних даних дозволяє не лише збільшити кількість описів, але й розширити їхній

семантичний діапазон за рахунок використання даних різної природи. Крім того, застосування розвинених інструментальних засобів, які надають нам сучасні інформаційні технології, створює можливість використовувати не лише набори даних різної структури, але й такі, що взагалі не мають структури.

Складність, однак, полягає в тому, що опис, створений людиною, за своєю природою є неформальним. Кожен автор вкладає в нього своє власне розуміння, яке не зводиться до дискретного набору уніфікованих термінів, послідовності яких легко порівнювати. Порівнюючи два описи, ми змушені маніпулювати не впорядкованими наборами фіксованих синтаксичних конструкцій, а їх змістом. Саме в побудові зв'язку між формальною множиною слів та їх змістом і виникає суб'єктивізм. Причина полягає в тому, що кожна людина, взагалі кажучи, будує такі зв'язки по-своєму, і відповідно по-своєму сприймає подібні зв'язки, побудовані іншими. Можна також сказати, що насправді ми матимемо справу з порівнянням не елементів реальності, а їхніх образів. Яскравим прикладом може служити проблема визначення необхідної міри оборони. Всі правильно усвідомлюють, що саме вчинив підсудний і не мають щодо цього розбіжностей. Але, залежно від суб'єктивної оцінки ступеню небезпеки, хтось вважає, що підсудний діяв в межах необхідної оборони, а хтось – що підсудний вийшов за її межі.

Отже, при застосуванні таких механізмів нам доведеться мати справу саме з неформалізованими задачами.

Ми розглянули лише один аспект проблеми, але він, на наш погляд, достатньо повно ілюструє загальний стан речей.

У випадку зловживань основні механізми мають ту ж саму природу, а фундаментальна відмінність полягає в тому, що

вони здійснюються свідомо з певною метою. Тому для зловживань сказане вище має той же самий сенс.

Але елемент цілеспрямованості створює додаткові теми для обговорення. Дійсно, мета особи, що вдається до зловживання, також несе в собі цілком конкретні відомості, іншими словами містить певну додаткову інформацію. Ми не знаємо цю мету, але можемо визначити її основні контури за семантичними кореляціями різних текстових наборів. Тут можливі дві групи методик, які за певних умов можуть поєднуватися.

До першої групи належать методики, які спираються на припущення щодо можливих зловживань певного характеру. Наприклад, при розгляді скарг не незаконне будівництво можна припустити, що зацікавлена сторона в тих чи інших експертних заключеннях свідомо ігноруватиме порушення забудовником ряду норм. Таким чином, ми можемо аналізувати наявні дані під відповідним кутом зору, що значно підвищує ефективність роботи.

Друга категорія містить методики, які навпаки, за специфічними характеристиками текстів можуть надати основу для припущення щодо наявності цілеспрямованого формування певної картини подій. Наприклад, порівнюючи склад наведених в заключенні характеристик ґрунтів з повним їх переліком (передбаченим законом) можемо виявити відсутність деяких із них. А це дає підставу припустити, що дані були сформовані неповними тому, що відсутні характеристики перешкождали будівництву. Звичайно, наведений приклад є гранично спрощеним, але він ілюструє сам підхід.

Зрозуміло, що для цього також треба здійснювати нетривіальні операції з довільними текстами, що неодмінно містять в собі неформалізовані задачі.

Слід розрізняти окремі сфери правової практики, в яких мають (або можуть мати) місце зазначені вище чинники. Як основні виділимо такі:

- практична діяльність фізичних та юридичних осіб, пов'язана з правовими проблемами;
- створення і використання нормативно-правової інформації в широкому розумінні.

Нас цікавитиме другий випадок.

Внесок цих чинників в інформаційні процеси в праві головним чином полягає в тому, що прийняття рішень може спиратись на хибну інформаційну основу, що в свою чергу приводить до негативних результатів. Підкреслимо, що дії тих осіб, які приймають рішення, можуть бути цілком адекватними, компетентними та добросовісними.

Важливу проблему становить ризик отримання помилкових результатів в процесах машинної обробки інформації. Дійсно, машина на кожному етапі своєї роботи використовує наявну інформацію відповідно до програмних інструкцій, причому вважається (на цьому засновані всі інформаційні технології), що дані є правильними. Якщо принаймні один інформаційний блок є хибним, машина буде працювати, нарощуючи кількість помилок. І, на відміну від людини, вона не зможе запідозрити наявність помилок.

Тому виявити дійсну причину помилковості рішення може бути доволі складно, незалежно від того, в який спосіб воно отримано.

Традиційно суб'єктивізм та зловживання виявляють і опрацьовують у окремих конкретних випадках, коли виникає відповідна підозра. Але сучасні інформаційні технології дозволяють організувати контроль великих масивів правової інформації. Він може здійснюватися як разова процедура



перевірки того чи іншого інформаційного масиву, або як постійний процес підтримання контенту на належному рівні.

Сам по собі такий контроль є складним процесом (в першу чергу внаслідок відсутності належної формалізації задач, що виникають при його здійсненні). Але він до того ж надзвичайно ускладнюється через великий обсяг правової інформації різного рівня структурованості. Тому сьогодні його слід розглядати скоріше як віддалену перспективу.

Застосування штучного інтелекту та машинної обробки даних в праві породжує ряд проблем, пов'язаних з контролем за їхнім використанням. Основні проблеми полягають у наступному:

- Інтерпретація результатів: машинне навчання та алгоритми штучного інтелекту можуть давати непередбачувані результати, які важко інтерпретувати та визначити, чи вони відповідають нормам та стандартам правової системи.
- Дискримінація: системи машинного навчання можуть навчитися дискримінувати на основі певних характеристик, таких як раса, стать, національність тощо, що порушує принципи рівності та недискримінації в правовій системі.
- Конфіденційність та приватність: машинна обробка даних може збирати та обробляти великі обсяги персональних даних, що ставить питання про захист конфіденційності та приватності цих даних.
- Відповідальність: відсутність чіткої відповідальності за результати та використання систем машинного навчання та штучного інтелекту може призвести до небезпечних ситуацій, які стосуються прав та безпеки людей.

- Етичні питання: використання систем машинного навчання та штучного інтелекту може відображати особисті погляди або норми тих, хто розробляє ці системи, що може порушувати принципи справедливості та рівності в правовій системі. Також можуть виникати питання, зокрема пов'язані з дискримінацією та безпекою даних.

Застосування машинного навчання в праві може бути проблематичним з кількох причин:

- Якість даних: якість даних є ключовим фактором для успішного використання машинного навчання. Однак, в правовій сфері документи можуть бути дуже складними для розуміння і інтерпретації, що може призвести до низької якості даних.
- Недостатньо даних: для успішного використання машинного навчання потрібна велика кількість даних. В правовій сфері не завжди можна зібрати достатню кількість даних, особливо якщо мова йде про нові ситуації або рідкісні події.
- Невідомість причинно-наслідкових зв'язків: машинне навчання може давати точні результати, але не завжди зрозуміло, яким чином вони були отримані. Це може стати проблемою в правовій сфері, де потрібно розуміти причинно-наслідкові зв'язки, щоб приймати обґрунтовані рішення.
- Відсутність регулятивного фреймворку: наразі відсутні чіткі правила та стандарти для застосування машинного навчання в праві, що може стати проблемою для його ефективного та безпечного використання.

Однією з проблем застосування систем машинного навчання та штучного інтелекту в праві є відсутність відповідальності за їх результати.

Оскільки алгоритми машинного навчання можуть бути дуже складними, важко передбачити, які конкретні результати будуть вироблені, тому стає проблемою встановлення відповідальності у випадку помилок або некоректної поведінки системи. Крім того, важко визначити, хто має нести відповідальність у випадку, коли помилка виникає через недостатність або некоректність вхідних даних, на яких навчалася система.

Іншою проблемою є те, що алгоритми машинного навчання можуть виробляти результати, які базуються на недосяжних для людини параметрах. Це може створити складнощі при визначенні причин виникнення певного результату, що у свою чергу може призвести до несправедливих рішень.

Крім того, системи машинного навчання можуть бути підвернені атакам з боку зловмисників, які можуть спрямовувати систему на небажані дії або впливати на результати. Такі атаки можуть бути складними для виявлення та запобігання.

Отже, необхідно розробляти відповідні правові механізми, які б дозволяли встановлювати відповідальність за результати застосування систем машинного навчання та штучного інтелекту. Також важливо забезпечувати надійність та захищеність систем від атак з боку зловмисників.

***Питання для самоконтролю:***

- Які негативні чинники відіграють суттєву роль у праві?
- Що таке суб'єктивізм і чому його слід вважати негативним чинником?

- Які можливі наслідки дій осіб та організацій, що припускаються суб'єктивізму?
- Чи можливо боротися з суб'єктивізмом і які механізми можна використовувати для його компенсації?
- Яка роль штучного інтелекту в парламентському контролі і чому для її реалізації потрібно належне навчання та компетентність фахівців?
- Як ефективно використання великих обсягів різномірних даних допомагає розширити семантичний діапазон описів?
- Які розвинені інструментальні засоби дозволяють використовувати набори даних різної структури або без структури?
- Як суб'єктивізм впливає на порівняння двох описів?
- Чому порівняння образів елементів реальності може бути проблемою?
- Які проблеми виникають у правовій системі через застосування штучного інтелекту та машинної обробки даних?
- Як машинне навчання та алгоритми штучного інтелекту можуть впливати на інтерпретацію результатів в правовій системі?
- Які характеристики можуть бути використані системами машинного навчання для дискримінації, та як це порушує принципи рівності та недискримінації в правовій системі?
- Як машинна обробка даних може вплинути на конфіденційність та приватність персональних даних в правовій системі?
- Хто несе відповідальність за результати та використання систем машинного навчання та штучного інтелекту, та як відсутність чіткої відповідальності може призвести до небезпечних ситуацій в правовій системі?

- Як використання систем машинного навчання та штучного інтелекту може порушувати принципи справедливості та рівності в правовій системі, та які етичні питання виникають у зв'язку з цим?
- Які можуть бути наслідки вироблення результатів, які базуються на недосяжних для людини параметрах?
- Які можуть бути наслідки атак з боку зловмисників на системи машинного навчання та штучного інтелекту?
- Які правові механізми можуть допомогти вирішити проблему відповідальності за результати застосування систем машинного навчання та штучного інтелекту?

## ВИСНОВКИ

У цьому посібнику розглянуто поняття «парламентській контроль» та нормативно-правова база його здійснення. Наведені важливі аспекти застосування інформаційних технологій у контексті парламентського контролю.

Надоно різноманітні приклади використання інформаційних технологій у парламентському контролі та детальний порівняльний аналіз різних методів та інструментів, що використовуються в цій галузі.

Підкреслено роль інформаційних технологій в забезпеченні збору, аналізу та доступу до правової інформації, обробці великих даних у галузі права, створенні мережевих структур, семантичних мереж та онтологій.

Також наведені застосування корпусної лінгвістики та порівняльного аналізу даних для підвищення ефективності парламентського контролю. Висвітлено можливості штучного інтелекту в процесі парламентського контролю та його роль у розвитку правової інформатики.

Проаналізовано різні типи мереж, включаючи мережі мови та семантичні мережі, показано їх важливість у відображенні взаємозв'язків між поняттями та термінами, які використовуються в правовій науці та практиці.

Показано можливості порівняльного аналізу даних, який дозволяє парламенту зрозуміти, як ефективно виконується його функція контролю, порівнюючи різні періоди часу та різні урядові органи.

Нарешті, висвітлено можливості застосування штучного інтелекту в парламентському контролі, наведені переваги і недоліки такого підходу. Штучний інтелект може допомогти в розробці систем автоматичного аналізу правової інформації, що значно полегшить роботу законодавчих органів.

Застосування інформаційних технологій у парламентському контролі є важливим кроком у напрямку ефективнішого виконання функцій законодавчої влади і перспективним напрямком, який може допомогти у вирішенні складних завдань, пов'язаних з забезпеченням правової діяльності законодавчих органів.

### **Рекомендована література до глав 2-6**

1. Електронна держава: нова ефективність урядування : монографія / А. А. Барікова. – Київ: Юрінком Інтер, 2016. – 224 с.
2. Міжнародна інформація. Кудрявцева С.П., Колос В.В., Навчальний посібник. – Київ: Видавничий Дім «Слово», 2005. – 400с.
3. Оброблення надвеликих масивів даних (Big Data) : навчальний посібник. / Д.В. Ланде, І.Ю. Субач, А.Я. Гладун. – Київ, ІСЗЗІ КПІ ім. Ігоря Сікорського, 2021. – 168 с.
4. Сучасне загальноєвропейське кримінальне законодавство: проблеми гармонізації : монографія / М. І. Хавронюк. – Київ: Істина, 2005. – 264 с.
5. Ланде Д.В. Елементи комп'ютерної лінгвістики в правовій інформатиці. – К.: НДІП НАПрН України, 2014. – 154 с.
6. Широков В.А., Бугаков О.В., Грязнухіна Т.О. Корпусна Лінгвістика. – Київ: Довіра, 2005. – 471 с.
7. Широков В.А. Елементи лексикографії.– Київ: Довіра, 2005. – 304 с.

Наукове видання

Дмитро Володимирович ЛАНДЕ  
Володимир Миколайович ФУРАШЕВ  
Сергій Михайлович БРАЙЧЕВСЬКИЙ

# **ІНФОРМАТИКА ПАРЛАМЕНТСЬКОГО КОНТРОЛЮ**

**Посібник**

В авторській редакції

Підписано до друку 15.12.2022. Формат 60x84/16. /16. Папір офс.  
Гарнітура Times. Спосіб друку – ризографія. Ум. друк. арк.  
14,88. Обл.-вид. арк. 9,29. Наклад 200 пр. Зам. № 10-160.

---

ТОВ "Інжиніринг"

ISBN 978-966-2344-80-6