

УДК 340:004.7:001.8

ЛАНДЕ Д.В., доктор технічних наук, старший науковий співробітник,
НДІ інформатики і права НАПрН України
ФУРАШЕВ В.М., кандидат технічних наук, доцент, старший науковий
співробітник, НДІ інформатики і права НАПрН України

ТЕРМІНОЛОГІЧНА МЕРЕЖЕВА МОДЕЛЬ ЯК ВІДОБРАЖЕННЯ ПРОЦЕСІВ ДЕНЦЕНТРАЛІЗАЦІЇ ВЛАДИ В УКРАЇНІ

Анотація. У роботі наводиться алгоритм побудови моделей термінологічних мереж з предметної області “децентралізація” на базі автоматизованого моніторингу і аналізу контенту мережі Інтернет. Терміни природної мови, що відображають поняття, склад і зміст яких динамічно змінюється, утворюють мережі. Дослідження цих мереж у динаміці дозволяє визначати нові аспекти у процесах децентралізації, спостерігати зміну пріоритетів, визначати напрямки соціально-правової активності. На прикладах показана адекватність підходу, що пропонується, а також, що кластери в термінологічних мережах можуть розглядатися як основа для визначення окремих наукових напрямків.

Ключові слова: модель предметної області, термінологічна база, контент-моніторинг, зв’язки понять, децентралізація.

Аннотация. В работе приводится алгоритм построения моделей терминологических сетей в предметной области “децентрализация” на базе автоматизированного мониторинга и анализа контента сети Интернет. Термины естественного языка, отражающие понятия, состав и содержание которых динамично изменяется, образуют сети. Исследование этих сетей в динамике позволяет определять новые аспекты в процессах децентрализации, наблюдать изменение приоритетов, определять направления социально-правовой активности. На примерах показана адекватность предлагаемого подхода, а также то, что кластеры в терминологических сетях могут рассматриваться как основа для определения отдельных научных направлений.

Ключевые слова: модель предметной области, терминологическая база, контент-мониторинг, связи понятий, децентрализация.

Summary. The algorithm of creation of models of subject domains on decentralization on the basis of computer aided data analysis and Internet content monitoring is offered in the article. Terms of natural language reflecting concepts, composition and content of which change dynamically, form networks. The study of these networks in dynamics allows us to define new aspects in the processes of decentralization, observe changes in priorities, and determine the directions of social and legal activity. Adequacy of the approaches offered is shown by example, and also the fact that clusters of terminological networks can be considered as a basis for detection of certain scientific directions.

Keywords: subject domain model, terminology-oriented database, content monitoring, links of concepts, decentralization.

Постановка проблеми. Децентралізація влади – це процес перерозподілу або диспергування функцій, повноважень, людей або речей від центрального управління [1]. Децентралізація влади (далі – децентралізація) здійснюється шляхом передання повноважень на прийняття рішень з головного органу будь-якої галузі уряду до органів нижчих рівнів. Тобто при цьому частина функцій центральної влади переходить до органів самоврядування; скасування або послаблення централізації. Цей процес який було названо “новим державним управлінням”, описаний як децентралізація, предметне управління, конкуренція урядової і місцевої координації [2].

Відповідно до Європейської Хартії Місцевого Самоврядування від 15 жовтня 1985 року., зміст місцевого самоврядування полягає у гарантованому державою праві та реальній здатності самих територіальних спільнот громадян (територіальних колективів) та сформованих ними органів самостійно та під свою відповідальність вирішувати окрему частину публічних справ, діючи в межах конституції та законів відповідної держави. Органи місцевого самоврядування визнано однією з головних підвалин будь-якого демократичного режиму.

Децентралізація в Україні має давню історію, що була започаткована ще у Конституції Української Народної Республіки (1918 р.). На цей час в Україні здійснюються процеси децентралізації – передачі значних повноважень та бюджетів від державних органів органам місцевого самоврядування. Вважається, що необхідно, щоб більше повноважень мали ті органи, які ближче до людей, де такі повноваження реалізуються успішніше.

Децентралізація являє собою динамічний процес, що охоплює всі ланки суспільно-правового поля країни. На різних етапах цього процесу застосовуються різні підходи, іноді правильні, іноді хибні. На цей час актуальною є проблема дослідження цих підходів, побудова моделей, що дозволяють вивчати процеси децентралізації, створення інструментальних засобів для підтримки прийняття рішень щодо процесів децентралізації особами, що приймають ці рішення.

Метою статті є опис термінологічної мережевої моделі на базі моніторингу мережевих ЗМІ, що відображає динаміку процесів децентралізації в Україні.

Виклад основного матеріалу. Як вважаємо, зазначена вище модель має застосовуватися як інструментальна база для підтримки прийняття рішень щодо процесів децентралізації особами, що приймають ці рішення. На цей час на веб-сайтах та у соціальних мережах в Інтернеті публікуються сотні тисяч повідомлень, що репрезентативно відображають стан соціально-правових процесів. На базі цієї інформації можлива побудова термінологічних мереж, що відбивають найважливіші поняття з предметних областей і семантичні зв'язки між ними [3]. За допомогою побудови і подальшого дослідження цих мереж можна визначати найбільш важливі і актуальні поняття, а також зв'язки між ними, динаміку їх застосувань у часі. Виявлення груп найбільш близьких за деякими змістовими характеристиками понять (кластерів) дозволяє визначати нові принципи класифікації процесів, що досліджуються, спостерігати зміну пріоритетів, визначати напрямки соціально-правової активності. Побудовані термінологічні мережі можуть бути розвинуті до рівня сучасних онтологій [4], когнітивних карт [5], що можуть ставати основою сценарного аналізу, відпрацювання соціально-правових процесів в предметній області (децентралізації), сюжетів майбутніх змін.

Методика побудови моделі.

На першому етапі побудови термінологічної моделі процесів децентралізації необхідно сформулювати репрезентативний текстовий корпус. Для цього авторами було застосовано систему контент-моніторингу InfoStream [6], за допомогою якої за запитом: *(децентралізац|деконцентрац|диспергуванн|делегуван|діволюц)&(влад)|(державн~управл інн)|самоврядуван)* було отримано 457 документів з україномовних веб-сайтів мережі Інтернет за визначеною тематикою за 1 – 12 червня 2017 року (Рис. 1).

На другому етапі здійснюється екстрагування термінів (окремих слів – уніграм, пари слів – біграм, і трійки – триграм) із відібраного на першому етапі текстового корпусу. Для підвищення точності термінів (міри їх відповідності тематиці децентралізації) до розгляду із текстового корпусу розглядаються лише абзаци, що відповідають первинному запиту.

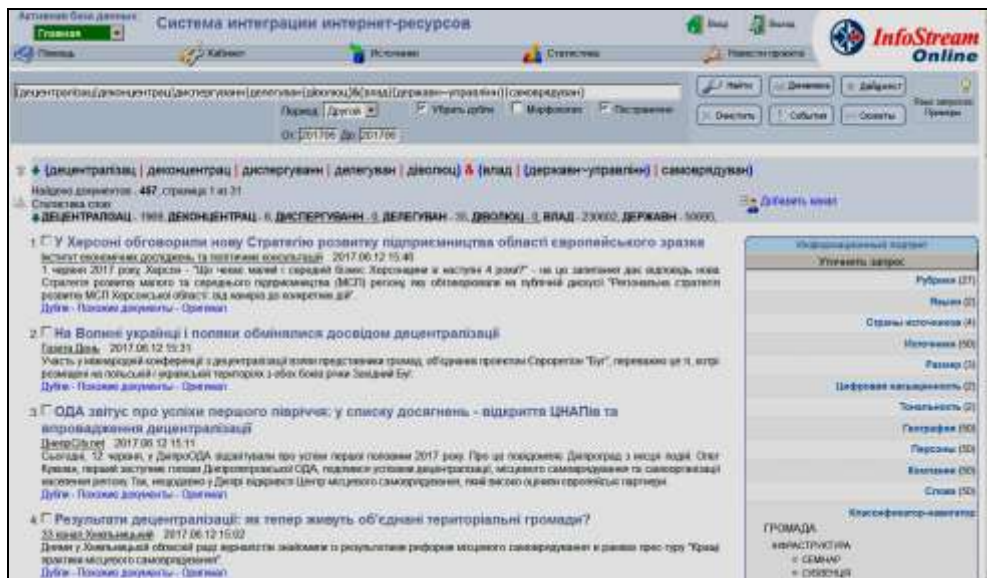


Рис.1 – Інтерфейс системи контент-моніторингу InfoStream

Витяг найбільш релевантних термінів здійснюється за дисперсійною моделлю [7]. Як міра важливості (ваги) термінів розглядається оцінка нерівномірності її входження у тексти. Для цього використовується така оцінка дискримінантної ваги терміна:

$$\sigma_i = \frac{\sqrt{\langle d^2 \rangle - \langle d \rangle^2}}{\langle d \rangle}$$

де: $\langle d \rangle$ – середнє значення послідовності d_1, d_2, \dots, d_n , n – кількість появ терміна t_i в інформаційному масиві. Тут d_k – відстань між появами терміна в тексті. Якщо позначити координати (номери) входження слова t_i в інформаційний масив як e_1, e_2, \dots, e_n , то $d_k = e_k - e_{k-1}$ ($e_0 = 0$).

На цьому етапі кожному окремому слову, словосполученню із двох слів (біграм) та із трьох слів (триграм) з текстового корпусу, що аналізується, ставиться у відповідність дисперсійна оцінка σ_i . Після цього всі терміни (окремо уніграми, біграми і триграми) сортуються за вагою і вибирається визначена експертами кількість найбільш вагомих цих термінів. Крім того, експерти додатково здійснюють візуальну фільтрацію слів, вилучаючи “зайві”, явно не релевантні тематиці.

На третьому етапі із відібраних термінів формується термінологічна мережа. Відомо декілька підходів до побудови мереж з текстів і різні способи інтерпретації вузлів і зв’язків, що призводить, відповідно, до різних видів представлення таких мереж. Вузли можуть бути сполучені між собою, якщо відповідні їм слова стоять поряд у тексті, належать до одного речення або абзацу, сполучені синтаксично або семантично [8].

Термінологічна мережа.

Традиційно розрізняють чотири різновиди мереж мови (просторів):

1. *L*-простір. Зв’язуються сусідні слова (терміни), 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. *C*-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 $R = R_{\max}$. У цьому випадку *L*-простір перетворюється на *P*-простір, який і застосовується у цій роботі.

Традиційний *P*-простір не передбачає направленості зв'язків мережі, у нашому випадку необхідно розглядати зв'язки із напрямками. Для визначення напрямків застосовуємо дві гіпотези:

- якщо один термін входить до складу іншого, то напрямок зв'язку встановлюється від першого до другого (наприклад “децентралізація” → “децентралізація влади”);
- якщо два вузли зв'язані у *P*-просторі, і перший має більший ступінь у ньому, то зв'язок встановлюється у напрямку від першого до другого.

На Рис. 2 відображено центральний фрагмент термінологічної мережі за тематикою “Децентралізація”.

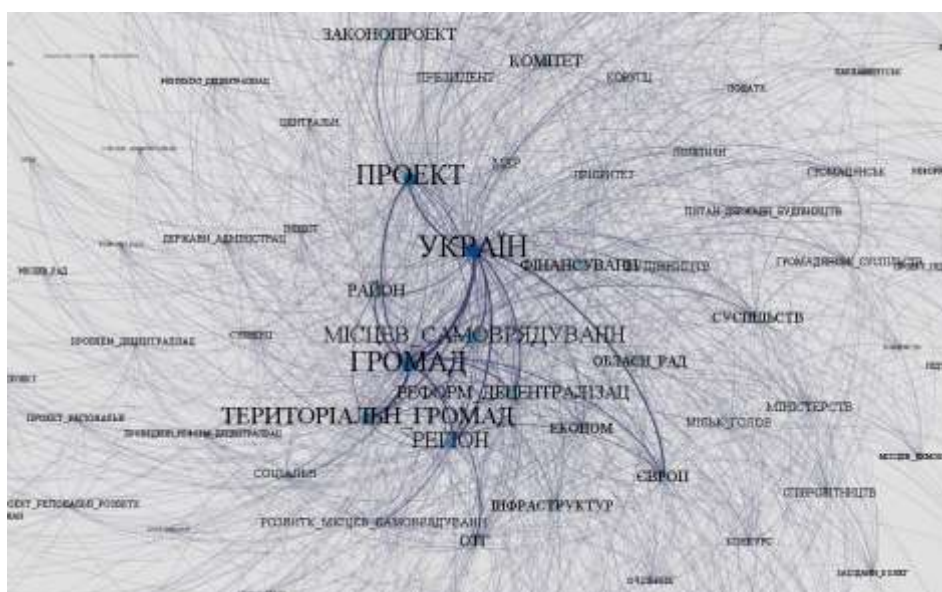


Рис. 2 – Фрагмент термінологічної мережі

Для виявлення подібних фрагментів мережі зазвичай застосовується кластерний аналіз, який дозволяє групувати близькі за деякими ознаками вузли (у нашому випадку – поняття). Для аналізу і візуалізації термінологічної мережі застосовується вільно розповсюджуваний програмний засіб Gephi (<http://gephi.org>). Цей засіб дозволив виявити найбільш вагомні поняття за мережевим критерієм PageRank (розроблений засновниками компанії Google). Результати кластеризації мережі наведено на Рис. 3.

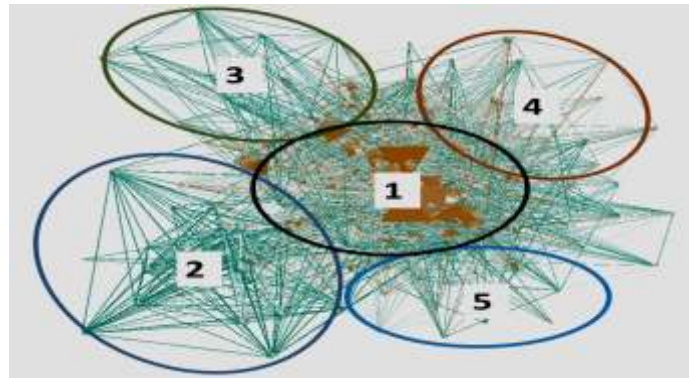


Рис. 3 – Кластерна структура мережі

Шляхом аналізу складу кластерів визначено такі основні напрямки термінології, що визначають процеси децентралізації в Україні на початок червня 2017 р.:

1	Загальні питання децентралізації	УКРАЇНА
		ДЕЦЕНТРАЛІЗАЦІЯ
		ПРОЕКТ
		ПРОЕКТ ДЕЦЕНТРАЛІЗАЦІЇ
		СТРАТЕГІЯ ОБ'ЄДНАННЯ
		ПРОЕКТ РЕГІОНАЛЬНОГО РОЗВИТКУ
		СУСПІЛЬСТВО
		ГРОМАДЯНСЬКЕ СУСПІЛЬСТВО
		ЗМІНИ КОНСТИТУЦІЇ УКРАЇНИ
2	Місцеве самоврядування	ГРОМАДА
		ГРОМАДИ УКРАЇНИ
		РОЗВИТОК МІСЦЕВОГО САМОВРЯДУВАННЯ
		ОТГ
		ОЧІЛЬНИК
		МЕР
		ФІНАНСУВАННЯ
		МІСЬКИЙ ГОЛОВА
3	Правове забезпечення	НОРМАТИВНО-ПРАВОВІ АКТИ
		ПРАВОВІ АКТИ
		АДМІНІСТРАТИВНО-ТЕРИТОРІАЛЬНА РЕФОРМА
4	Державні органи, посадові особи	ВЕРХОВНА РАДА
		КОМІТЕТ
		АГРАРНИЙ КОМІТЕТ
		ПРЕМ'ЄР ВОЛОДИМИР ГРОЙСМАН
		АНДРІЙ ПАРУБІЙ
5	Європейський вибір	ЄВРОПЕЙСЬКІ КРАЇНИ
		ОФІС РАДИ ЄВРОПИ
		ЄВРОПЕЙСЬКА ІНТЕГРАЦІЯ

Висновки.

Децентралізація – складний, багатогранний процес, який визначає не лише перерозподіл владних повноважень по всій ієрархії системи державного управління, а й перерозподіл сфер обов’язків та відповідальності, фінансових і інформаційно-правових потоків. В цих умовах, особливо з огляду на суттєву трансформацію суспільних відносин, дуже велику роль відіграють саме інформаційно-правові потоки, питання запобігання та своєчасного усунення реальних та потенційних колізій у правовому полі, що стрімко буде розширюватися.

Децентралізація являє собою динамічний процес, що охоплює всі ланки суспільно-правового поля країни. На різних етапах цього процесу застосовуються різні підходи, іноді правильні, іноді хибні.

У роботі запропоновано і реалізовано алгоритм формування моделей предметних областей шляхом автоматизованого аналізу контенту мережі Інтернет. Від статичних моделей предметних областей таких підхід відрізняється урахуванням динамічної зміни контенту бази даних цього сервісу, урахуванням нових понять, феноменів, що з’являються, зокрема в юридичній області, що розглядається.

Процеси децентралізації знаходять своє відображення у мовному середовищі. Терміни природної мови, що відображають поняття, склад і зміст яких динамічно змінюється, утворюють мережі. Дослідження цих мереж у динаміці дозволяє визначити нові аспекти у процесах децентралізації, спостерігати зміну пріоритетів, визначити напрямки соціально-правової активності.

Запропонована модель, що базується на моніторингу і подальшому аналізі мережевого контенту, має застосовуватися як інструментальна база для підтримки прийняття рішень щодо процесів децентралізації особами, що приймають ці рішення.

Використана література

1. Definition of decentralization, TheFreeDictionary.com, accessed February 4, 2013.
2. Managing Decentralisation : A New Role for Labour Market Policy, Organisation for Economic Co-operation and Development, Local Economic and Employment Development (Program), OECD Publishing, 2003. – P. 135.
3. Ландэ Д.В., Снарский А.А., Путятин В.Г. Построение терминологической сети предметной области // Реєстрація, зберігання і обробка даних. – 2014. – Т. 14. – № 2. – С. 114-121.
4. Ландэ Д.В., Снарский А.А. Автоматическое построение терминологических онтологий // Proceedings of the XX-th International Conference Knowledge-Dialogue-Solution 2014, Kyiv, Ukraine. – Pp. 28-71.
5. Кузнецов О.П. Когнитивное моделирование слабо структурированных ситуаций / О.П. Кузнецов. – М. : Политехнический музей, 2006. – С. 86-100.
6. Ланде Д.В. Программно-апаратний комплекс інформаційної підтримки прийняття рішень : науково-методичний посібник / Д.В. Ланде, В.М. Фурашев, О.М. Григор’єв. – К. : Іжініринг, 2006. – 48 с.
7. Ortuño M., Carpena P., Bernalda P., Muñoz E., Somoza A.M. Keyword detection in natural languages and DNA // Europhys. Lett., 2002. – 57. – P. 759-764.
8. Ланде Д.В. Элементы компьютерной лингвистики в правовой информатике. – К.: НДПП НАПрН України, 2014. – 168 с.

~~~~~ \* \* \* ~~~~~