

Інформатизація та інформаційні технології

УДК 004.67

БРАЙЧЕВСЬКИЙ С.М., кандидат фізико-математичних наук,
ЛАНДЕ Д.В., доктор технічних наук

ІНТЕРАКТИВНІ ГІПЕРТЕКСТОВІ ІНФОРМАЦІЙНІ РЕСУРСИ В ПРАВОВІЙ НАУЦІ

Анотація. *Можливості сучасних гіпертекстових систем дозволяють створювати потужні інформаційні системи. Запропоновано методика створення структури зв'язків гіпертекстової енциклопедії, що ґрунтується на алгоритмах визначення дискримінантної сили термінів у текстах і принципах побудови мережі природної ієрархії термінів.*

Ключові слова: електронна енциклопедія, гіпертекст, Інтернет, екстрагування понять, MediaWiki.

Аннотация. *Возможности современных гипертекстовых систем позволяют создавать мощные информационные системы. Предложена методика создания структуры связей гипертекстовой энциклопедии, которая основана на алгоритмах определения дискриминантной силы терминов в текстах и принципах построения сети естественной иерархии терминов.*

Ключевые слова: электронная энциклопедия, гипертекст, Интернет, экстрагирование понятий, MediaWiki.

Summary. *The possibilities of modern hypertext systems allow you to create powerful information system. The technique of creating a hypertext link structure encyclopedia proposed. The technique is based on the algorithm for determining the discriminant power of the terms in the texts and the principles of building a natural hierarchies of terms networks.*

Keywords: electronic encyclopedia, hypertext, Internet, feature extraction, MediaWiki.

Постановка проблеми. Електронні енциклопедії, в першу чергу “Вікіпедія” в усіх її редакціях, за останні роки набули значної популярності в сучасних мережних ресурсах. Один з важливих моментів полягає в тому, що такі ресурси створюють надзвичайно широкі можливості для формування користувачами мережі Інтернет власного контенту. Головне при цьому полягає у тому (і це виявилось можливим), що – електронні енциклопедії успішно заповнюються статтями високого професійного рівня, які пишуть ентузіасти з власної ініціативи.

Поширення електронних енциклопедій значною мірою зумовлено наявністю технології MediaWiki, що, як відзначалося, дозволяє користувачам мережі Інтернет самостійно створювати інформаційні ресурси відповідно до власних потреб. А ця обставина робить відповідні технології доступними для широкого кола користувачів. Тому певний інтерес становить дослідження таких можливостей.

Метою статті є викладення деяких результатів досліджень щодо формування гіпертекстових електронних енциклопедій на базі інформаційних ресурсів, як наявних, так і тих, що створюються користувачами спеціально для цієї мети.

Виклад основних положень. Гіпертекст є одним з важливих феноменів ХХ-го ст., що заслуговує на глибоке вивчення в контексті переходу до інформаційного суспільства. Цей термін запровадив в 1965 р. Т. Нельсон [1] для опису непослідовного запису. Йшлося щодо нелінійної організації тексту, яка містить в собі окремі фрагменти

в довільній послідовності, та набір посилань, за допомогою яких можна переходити від одного фрагменту до іншого.

Слід зауважити, що нелінійна побудова текстів використовувалася досить давно [2], і загальновідомим прикладом може служити Біблія. Але саме Нельсон вперше поставив задачу свідомого формування нелінійної організації текстів для досягнення певних цілей і започаткував відповідну технологію.

Сучасне розуміння гіпертексту суттєво розширилося включенням до його структури таких нетекстових в традиційному розумінні компонентів як зображення, аудіо- та відеоматеріали [3].

Переваги гіпертексту.

Вважається, що головною перевагою гіпертекстової моделі є те, що вона цілком відповідає асоціативному характеру генерації та обробки інформації людською свідомістю [2]. Тому вона претендує на роль структури для ефективного представлення та передачі знань [4]. Дійсно, нехай в лінійній моделі тексту маємо одиницю смислу, з якою пов'язано декілька інших. Очевидно, що для того, щоб знайти одну із них, треба прочитати весь фрагмент тексту, що розташовано між ними, а він, взагалі кажучи, може бути досить великим. До того ж, деякі з фрагментів можуть бути розташовані в зворотному порядку. В гіпертекстовій моделі всі одиниці змісту, пов'язані з даною, можуть бути отримані безпосередньо за допомогою прямих посилань. Вони утворюють мережу (її ребра представляють зв'язки між окремими фрагментами тексту, що мають визначений зміст), яка раціонально відбиває структуру сприйняття складних ідей людською свідомістю. Це, в свою чергу, призводить до зростання адекватності сприйняття загального змісту споживачем. Гіпертекст є реальною технологією реалізації загального принципу, сформульованого В. Бушем [5] ще у 1945 році: кожний елемент інформації має бути відправною точкою для безпосереднього автоматичного вибору користувачем будь-якого іншого елемента інформації.

Оскільки в базах знань ключову роль відіграють саме зв'язки між елементами даних, гіпертекстова модель є однією з найперспективніших в плані розвитку цієї галузі. А тому проблема вивчення гіпертексту та створення технологій для його генерації та використання зберігає свою актуальність.

Гіпертекст і правова наука.

Однією з галузей, в яких гіпертекст забезпечує оптимальну форму організації даних, є, безперечно, правова наука. Перш за все тому, що право оперує величезною кількістю окремих інформаційних одиниць, формально не пов'язаних між собою. Дійсно, маємо складну та розгалужену систему інформаційних розділів, структура якої описується багаторівневою таксономією, що має цілу низку ієрархічних ланцюжків, таксони яких частково перетинаються. Також слід мати на увазі, що різні школи правової науки, які відрізняються типами праворозуміння, мають різні відповідні класифікаційні схеми. А це суттєво збільшує обсяг даних.

Проілюструємо цю ситуацію на конкретному прикладі, взятому з української “Вікіпедії”. Наведені матеріали можуть відрізнитись від інших джерел, але у даному випадку розглядається лише структура даних (розбіжність джерел лише підкреслює складність тексту, що містить правову інформацію).

Одним із найбільш загальних є поняття “система права” [6], яку не слід плутати з поняттями “правова система” і “система законодавства”. Система права – це сукупність чинних принципів і норм права.

Відомості, пов’язані з цим поняттям, на загальному рівні розподіляються на такі розділи:

1. Загальні ознаки норм права.
2. Види норм права.
3. Спеціалізовані норми права.
4. Генезис норм права в романо-германській правовій сім’ї.
5. Структура норми права.
6. Джерела.

Кожний з них має власну таксономію нижчого рівня. Так, наприклад, до загальних ознак норм права відносяться наступні підрозділи:

1. Правило поведінки регулятивного характеру.
2. Загальнообов’язкове правило поведінки.
3. Правило поведінки загального характеру.
4. Формально-визначене правило поведінки представницько-зобов’язуючого змісту.
5. Правило поведінки, прийняте в суворо встановленому порядку.
6. Правило поведінки, забезпечене всіма заходами державного впливу, аж до примусу.

Норми права розподіляються на класичні та спеціалізовані. Далі, вони класифікуються:

1. За предметом правового регулювання (норми конституційного права, адміністративного права, кримінального права тощо).
2. За методом правового регулювання (імперативні та диспозитивні).
3. За характером впливу на особу (заохочувальні та рекомендаційні).
4. По субординації в правовому регулюванні (матеріальні та процесуальні).
5. За ступенем визначеності варіанта поведінки (абсолютно визначені та відносно визначені).

І так далі. Вже на цьому надзвичайно простому прикладі добре видно, що текст, який описує дану тему, неможливо представити в лінійному вигляді. Дійсно, практично неможливо помістити в наведені текстові дані відомості, скажімо, про конституційне право, адміністративне право, тощо, не змінивши суттєво їх структуру. І зрозуміло, що такі зміни структури тексту, здійснені у повному обсязі, зроблять його вкрай незручним для засвоєння.

Як правило, вважається, що гіпертекст створюється свідомо і цілеспрямовано, з певною, заздалегідь визначеною метою. Такий процес формування контенту є доволі досконалим з точки зору забезпечення його цілісності та структурної довершеності. Але він надто обмежений можливостями автора (або нечисленної групи авторів, в якій можливе повне узгодження дій окремих учасників), в тому числі щодо визначення актуальних тем. Автор гіпертексту може просто не знати, що з певним елементом даних пов’язані деякі цікаві та важливі речі. Тому стає очевидним, що участь широкого кола учасників формування контенту суттєво підвищує його рівень. Для цього, в свою чергу, потрібні інструментальні засоби, які дозволяють користувачу системи доповнити вже існуючий гіпертекст (редагування також є важливим компонентом, але в ньому немає нічого специфічного для гіпертексту, і ми його не розглядатимемо). Ці інструментальні засоби повинні забезпечувати виконання двох видів операцій над гіпертекстом:

- додавання до існуючого гіпертексту одного або більше фрагментів;
- виділення в існуючому гіпертексті відповідних слів або наборів слів і пов’язування з ними посилання на додані фрагменти гіпертексту.

Гіпертекст – система, що саморозвивається.

Важливим є те, що дані операції мають бути повністю уніфікованими, щоб з гіпертекстом могли ефективно працювати різні користувачі, а також щоб комп'ютерні програми могла правильно маніпулювати ними.

Такі інструментальні засоби неважко розробити, але проблема уніфікації їх набагато складніша. Головна причина полягає в тому, що сучасні глобальні мережі не регламентовані жорсткими стандартами, що сприяє їх технологічній різноманітності. В тому числі, відсутні належні стандарти, які визнавалися б “мережною спільнотою” як обов'язкові. Рішенням проблеми може бути, наприклад, уніфікація процесу роботи з даним конкретним гіпертекстом, що забезпечується функціональними можливостями програмного комплексу, призначеного для його генерації та супроводу. Саме так і працюють електронні енциклопедії, засновані на технології MediaWiki, яка є основою побудови веб-систем, призначених для колективної розробки, зберігання, структуризації тексту, гіпертексту, файлів, у тому числі мультимедіа. Усі сторінки MediaWiki-сайту є статтями, вміст яких представляє з себе текст, в якому можна використовувати просту Wiki-розмітку або теги HTML.

Саме технологію MediaWiki було вибрано як базову для створення Електронної енциклопедії законодавства України виходячи з таких 7 головних причин [6]:

1. З самого початку мережева технологія, успішність підтверджена практикою.
2. Вільна ліцензія, не потрібна власна розробка програмного забезпечення.
3. Можливість включення великої кількості інформаційних адміністраторів.
4. Можливість організації гіперпосилань як на зовнішні ресурси, так і на окремі свої статті.
5. Можливість створення різних категорій нарощування видів інформаційних ресурсів.
6. Можливість “підключення” мультимедійних матеріалів.
7. Можливість пакетного завантаження вже накопичених матеріалів.

Головною перевагою технології MediaWiki є, безперечно, комплекс вбудованих інструментальних засобів, які дозволяють виконувати штатні операції з даними в інтерактивному режимі через стандартний веб-інтерфейс.

Даний підхід робить початкове проектування гіпертексту другорядною справою. Більше того, гіпертекст на момент створення програмно-апаратного комплексу може взагалі не проектуватись. Мають бути лише однозначно задані його характеристики. Це потрібно для того, щоб забезпечити відповідність кінцевого продукту початковій меті його створення. Між іншим, саме так була створена “Вікіпедія” – розробники лише створили технічні засоби для наповнення системи статтями. Склад та зміст статей віддавався на розсуд користувачам. І цей проект виявився цілком успішним.

Але він має свій зворотній бік – низький рівень оптимізації структури гіпертексту. Інформаційне наповнення системи широким колом незалежних одне від одного учасників цього колективного процесу робить гіпертекст інформаційно насиченим, але далеким від досконалості в плані його структури. Наприклад, це може приводити до наявності надлишкових зв'язків. Взагалі кажучи, розвинений гіпертекст може мати нетривіальну топологію, але це питання виходить за рамки даної роботи.

Мережева (графова) структура гіпертексту.

Активне використання мережних технологій, які завдяки використанню інтерактивних інструментальних засобів перетворюють формування гіпертексту в динамічний процес, який не має визначеного (навіть умовного) завершення. Необмежене в часі наповнення гіпертексту новими текстовими масивами, пов'язаними

між собою багаторівневими нелінійними зв'язками, робить продукт, що утворюється, надзвичайно складним.

Тому виникає актуальна задача вивчення таких нелінійних систем і розробка засобів навігації в гіпертексті невизначеної наперед структури.

Одним з найбільш ефективних методів опису (а отже, і вивчення) складних гіпертекстів є застосування теорії графів [8]. Цей метод цілком успішно практикується ще з 1980-х років як в теоретичному, так і прикладному планах [див. наприклад, 9, 10]. Важлива перевага цього методу полягає в тому, що теорія графів добре розроблена, і це дає можливість використовувати розвинений математичний апарат для дослідження довільного гіпертексту, абстрагуючись від його конкретного змісту.

В основі методу лежить просте правило відповідності: вузлам графу відповідають інформаційні одиниці, що утворюють гіпертекст, а ребрам – гіперпосилання, що реалізують зв'язки між ними. Типовий граф наведено на Рис. 1.

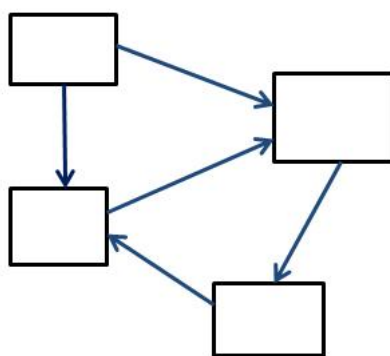


Рис. 1 – Простий приклад представлення гіпертексту графом.

Наступним кроком у вивченні гіпертекстів є перехід від простих графів до складних мереж [11]. Також актуальним завданням є оптимізація гіпертекстових мереж, у тому числі заміною надлишкових ланцюжків транзитивних посилань прямими посиланнями. Дослідження реальних гіпертекстових систем може здійснюватися за допомогою автоматизованого аналізу мережі Інтернет за допомогою спеціально розроблених методик і інструментальних засобів.

Побудова гіпертексту на основі послідовного масиву текстів.

На цей час можливості сучасних технологій (зокрема MediaWiki) дозволяють створювати гіпертекстові енциклопедичні ресурси (зокрема, їх первинне заповнення) на основі заздалегідь підготовлених документальних масивів. При цьому формування гіперпосилань здебільшого здійснюється експертами-аналітиками після завантаження документів у спеціально призначене гіпертекстове середовище. Саме цей етап роботи вважається найбільш відповідальним і тривалим. У зв'язку з цим авторами розроблена методика автоматичного формування гіперпосилань, що базується на алгоритмах автоматичного екстагування найвагоміших термінів предметної галузі – слів і словосполучень із текстових документів [12, 13]. На Рис. 2 наведено приклад мережі природніх ієрархій термінів [14], яку побудовано на основі масиву документів, які складають контентну основу моделі Електронної енциклопедії законодавства України.

Після виявлення найвагоміших термінів, вони визначаються як джерела для гіперпосилань, які ведуть до документів, до заголовків яких вони входять. Таким чином формується дієва система посилань, яка у сукупності визначає структуру гіперпосилань енциклопедії. Ці гіперпосилання виступають, з одного боку, у якості

“навігаторів” по матеріалах енциклопедії, а з іншого боку, базою для побудови графічних мап зв’язків, візуалізація яких дозволяє в узагальненому вигляді досліджувати систему понять у галузі права. Як приклад засобів для подібної візуалізації динамічної структури зв’язків понять в електронній енциклопедії можна навести систему WikiLinks, доступну, зокрема, користувачам мобільних застосунків компанії Apple.

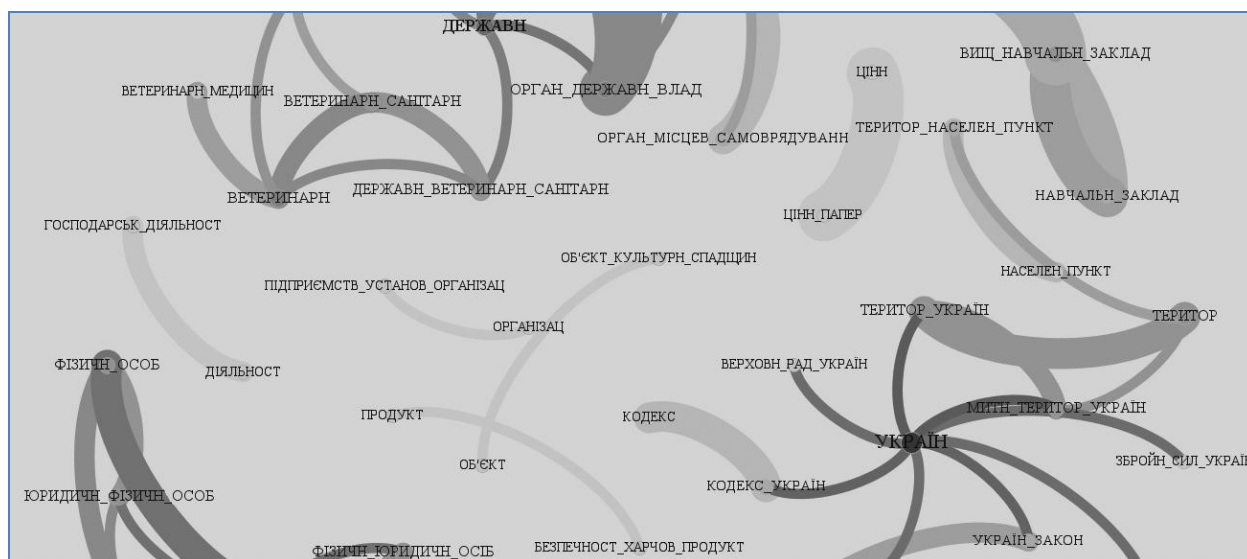


Рис. 2 – Фрагмент мережі природної ієрархії термінів.

Висновки.

1. Показано, що гіпертекст є дійсно ефективним засобом організації великих обсягів інформації. Важливим є те, що існують методи аналізу гіпертекстових систем, які дозволяють підвищувати рівень їх проектування та досягати значної оптимізації як структури, так і функціональних можливостей.

2. Можливості гіпертекстових систем значною мірою ґрунтуються на готових технологічних рішеннях, які дозволяють створювати різноманітні апаратно-програмні комплекси для накопичення та обробки інформації.

3. Технологія MediaWiki, що розроблена для створення електронних енциклопедій, насправді має значно ширші можливості, які дозволяють формувати складні гіпертекстові масиви майже без попереднього проектування, і при цьому забезпечувати цілісність кінцевого продукту та семантичну єдність його окремих фрагментів.

4. Запропоновано методика створення структури зв’язків гіпертекстової енциклопедії, що ґрунтується на алгоритмах визначення дискримінантної сили термінів у текстах і принципах побудови мережі природної ієрархії термінів.

Такі технології, безперечно, є перспективними в галузі правової науки, оскільки дозволяють створювати інформаційно-насичені системи, що органічно доповнюють традиційні бази даних.

Використана література

1. Nelson T.N. A file structure for the complex, the changing, and the indeterminate // ACM 20th National Conference Proceedings (Cleveland, Ohio, 1965). – P. 84-100.

2. Эпштейн В.Л. Введение в гипертекст и гипертекстовые системы. – Режим доступу : [//www.lingvolab.chat.ru/library/hypertext.htm](http://www.lingvolab.chat.ru/library/hypertext.htm)

3. John B. Smith, Stephen F. Weiss, and Gordon J. Ferguson. A Hypertext Writing Environment and Cognitive Basis. Proceedings of Hypertext'87. ACM Press, 1987.
4. Carlson P.A. A Way of Incorporating User Feedback into Online Documentation. In E.Barret (Ed) // Text, ConText, and Hypertext. Writing with and for the Computers. MIT Press series in information systems. – 1988. – 93-110. – P. 1-29.
5. Bush V. As We May Think //Atlantic Montly. – 1945. – № 7. – P. 101-108.
6. – Режим доступу : http://uk.wikipedia.org/wiki/Система_права
7. Ланде Д.В., Брайчевський С.М. Можливості довідкових мережевих ресурсів для створення електронної енциклопедії законодавства України // Інформація і право. – 2013. – № 2(38). – С. 72-76.
8. Берж К. Теория графов и ее приложения / К. Берж. – М. : ИЛ, 1962. – 320 с.
9. Trigg R.H. A Network-Based Approach to Text Handling for the Online Scientific Community / Ph.D. R.H. Trigg. – University of Maryland, 1983.
10. Conklin J. Hypertext : an introduction and survey // Computer. – 1987. – № 9.
11. Ландэ Д.В. Интернетика : навигация в сложных сетях: модели и алгоритмы / Д.В. Ландэ, А.А. Снарский , И.В. Безсуднов. – М. : Либроком (Editorial URSS), 2009. – 264 с.
12. Ланде Д.В. Елементи комп'ютерної лінгвістики в правовій інформатиці / Д.В. Ланде. – К. : НДІП НАПрН України, 2014. – 168 с.
13. Ланде Д.В. Методи оцінки рівня дискримінантної сили слів у текстах з правової тематики // Правова інформатика. – 2012. – № 3(35). – С. 5-9.
14. Ланде Д.В. Формування мереж природних ієрархій термінів на основі аналізу текстових корпусів з правової тематики // Правова інформатика. – 2014. – № 1(41). – С. 12-17.

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