

Олег Сербін,

Молодший науковий співробітник НБУВ

## **Представлення інформаційно-пошукових мов у веб-орієнтованих системах**

*Об'єктом даної публікації є інформаційно-пошукові мови (ІПМ) у веб-орієнтованих системах, а предметом – представлення ІПМ у веб-орієнтованих системах. Метою ж даної публікації є висвітлення представлення ІПМ у веб-орієнтованих системах.*

Інформаційні масиви, що є змістовними компонентами усього загалу Інтернет, на сьогоднішній день невпинно збільшують свою чисельність та масштабність відносно цільового призначення та застосування. З формуванням інтернет-ресурсів виникає закономірна потреба каталогізаційно формалізувати всю інформацію за для полегшення: а) розміщення; б) пошуку; в) використання. Тому імпортаційні процеси щодо представлення інформаційно-пошукових мов не тільки в традиційному<sup>1</sup>, а й інтерактивному просторі вже сьогодні цікаві з огляду на їх формну новизну та різноаспектність виконання. Цим пояснюється *актуальність* даної публікації.

У відповідності до міждержавного ГОСТу 7.74—96 "Информационно-поисковые языки. Термины и определения", ІПМ – це формалізована, штучна

---

<sup>1</sup>Питання ІМП в традиційній формі нами розглянуті в попередніх публікаціях: Сербін О. УДК українською мовою: історичні аспекти створення та аналітичний огляд структури / Олег Сербін // Бібл. вісн. – 2008. – № 5. – С. 10–16; Сербін О. Конгломерат інформаційно-пошукових мов як консолідаційна модель загального механізму впорядкування та пошуку бібліографічної інформації / Олег Сербін // Бібліотечний вісник. – 2008. – № 1. – С. 3–10 ; Сербін О. «Рубрикатор НБУВ»: історія започаткування та структурний аналіз / Олег Сербін // Вісн. Кн. палати. – 2008. – № 1. – С. 17–20; Сербін О. Використання бібліотечно-бібліографічних класифікацій в Україні : сучасний стан та тенденції розвитку / Олег Сербін // Наук. пр. Нац. б-ки України ім. В. І. Вернадського. – К., 2007. – № 19. – С. 77–83; Сербін О. Система десяткової класифікації українською мовою: історія та структурний аналіз / Олег Сербін // Вісн. Кн. палати. – 2007. – № 10. – С. 19–22; Сербін О. Основні десяткові класифікації українською мовою радянського періоду: історія та структурний аналіз / Олег Сербін // Вісн. Кн. палати. – 2007. – № 9. – С. 18–21; Сербін О. Бібліотечно-бібліографічна класифікація: не лише актуальна, а й перспективна / Олег Сербін // Вісн. Кн. палати. – 2007. – № 4. – С. 21–25; Сербін О. Міжнародна децимальна бібліографічна класифікація / Олег Сербін // Вісн. Кн. палати. – 2007. – № 8. – С. 19–22; Сербін О. Основні класифікаційні системи XIX – XX ст. / Олег Сербін // Вісн. Кн. палати. – 2006. – № 11. – С. 33–38.

мова, призначена для індексування документів, інформаційних запитів та описування фактів з метою подальшого зберігання та пошуку.

На нашу думку, типологічно ІПМ, що представлені у веб-орієнтованих системах можна розділити на чотири групи: *класифікаційні ІПМ; предметні ІПМ; дескрипторні ІПМ; фолксонемічні ІПМ.*

Класифікаційна ІПМ – це мова, що покликана виконувати свої основні функції за допомогою будь-якої класифікаційної системи.

Предметизаційна ІПМ – це мова, що покликана виконувати свої основні функції за допомогою певних предметних рубрик, певного зібрання останніх [1].

Дескрипторна ІПМ – це мова, що покликана виконувати свої основні функції за допомогою дескрипторів.

Фолксонемічна ІПМ – це мова, що покликана виконувати свої основні функції за допомогою будь-якої фолксонемічної онтології<sup>2</sup>.

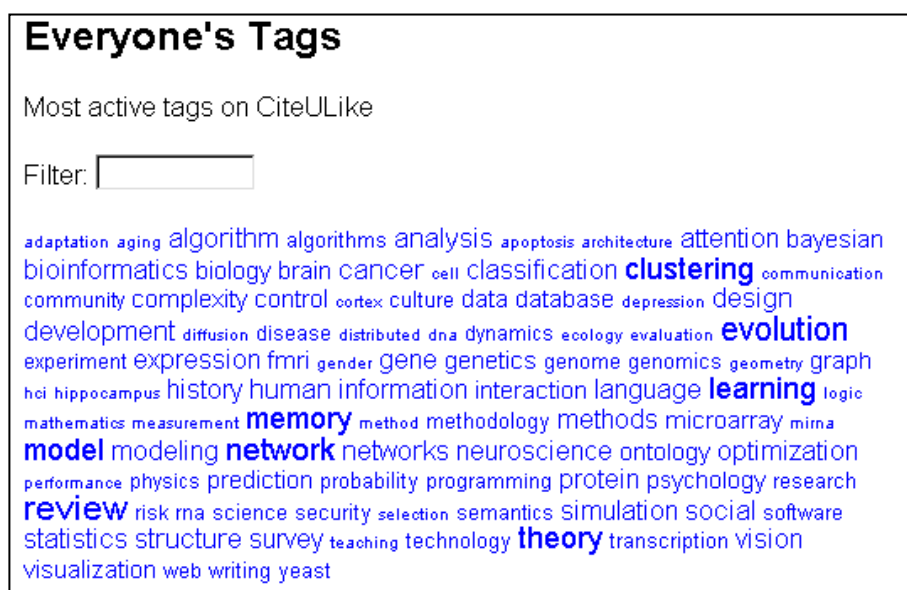
Відкорінне представлення фолксонемічних онтологій в середовищі інтернет зумовлює їх загальнодоступність та безмежну всеосяжність, стосовно аспектів представлення змісту інформаційних ресурсів, термінологічними координатами.

Так, як приклад, можна розглянути, інтернет-ресурс <http://www.citeulike.org/> [3] – зразок *колтегу* (колективного тегування) інформаційних ресурсів в інтернет середовищі. За допомогою цього ресурсу можливо здійснювати пошук потрібної інформації, формуючи пошуковий запит у відповідній формі інтерфейсу даного сайту. Крім того, після реєстрації користувачу надається можливість формувати статті щодо потрібної тематики, відносно відповідних ключових слів. Як наслідок реально можливо створювати користувачеві свою власну бібліотеку, в якій висвітлюються дані про користувача (фото, ім'я, місце роботи, посада), результати його дій на сайті (пошук, створення нових статей, кількість авторів та тегів).

---

<sup>2</sup> Онтологія – це спроба широкоформатної і детальної формалізації певної галузі знань за допомогою концептуальної схеми

Крім того, найбільш популярні ключові слова візуально представлено хмарою тегів, за допомогою якої наочно видно ступінь використання (пошуку) того чи іншого терміну.



Доцільною також є можливість експортувати бібліотеку в файл BibTeX – програму, що створена для форматування та роботи з бібліографічними списками, використовується разом з LaTeX<sup>3</sup>.

В процесі роботи на сайті постійно зустрічаються статті, що цікавлять користувача. Щоб уникнути масового відкривання кожної сторінки, що цікавить користувача, передбачено функцію формування списку. Формується список простим натисканням кнопки, і, таким чином, створюються певні закладки для кожної обраної сторінки, що в кінцевому результаті має форму переліку у хронологічній формі.

Таким чином можна бачити всі умови для роботи в віртуальних бібліотеках, що представлені на даному сайті. Пошук здійснюється фолксонемічно і поповнення бази тегів відбувається демократично та відкрито. Можна бачити багатоаспектність ключових слів відносно одного поняття чи статті. Продемонстровано повну свободу дій щодо формування електронної бібліотеки та відповідного списку тегів.

<sup>3</sup>Найпопулярніший набір макропакетів до TeX, що полегшує набір складних документів. Пакет дозволяє автоматизувати велику кількість різноманітних завдань набору тексту і підготовки статей, включаючи набір тексту на різних мовах, нумерацію розділів, перехресні посилання. – BibTeX // Wikipedia: the free encyclopedia. – Access Path: URL: <http://ru.wikipedia.org/wiki/BibTeX>

Що ж до класифікаційних ПІМ, що представлені в інтернет, то їх можна розглянути на прикладі інформаційно-пошукового ресурсу <http://bubl.ac.uk> [2], зміст в якому організовано за допомогою Десяткової класифікації Д'юї (DDC).

Взагалі, BUBL LINK Catalogue – це зібрання інтернет-ресурсів академічного спрямування. Інформацію тут впорядковано різноінструментально, незважаючи на однотипність представлення інформаційного пошуку на даному сайті. Так, можливим є пошук за ключовими словами (Search/Advanced Search). Варто зазначити, що пошуковий образ документа можна виражати як літерно, так і цифрами. Що правда, в процесі пошукової команди відбувається пошук цифр та їх комбінацій за назвами, заголовками та анатоційними даними документів. Індексне вираження рубрики, до якої документи відносяться, в розрахунок пошуку не приймається.

Окремо також проводиться пошук за допомогою *Subject Menus* – переліку предметів. Дане меню – не що інше, як перелік предметів посилкового характеру. Тобто, приміром, шукаючи по предмету *Finance*, ми автоматично отримуємо на екрані наступний список:

### **332 Financial economics**

See also [public finance](#)

[332 Financial economics: general resources](#)

[332.1 Banks](#)

[332.4 Money](#)

[332.6 Investment and investments](#)

[332.64 Exchange of securities and securities exchanges](#)

Зайшовши ж в рубрику, що найбільше нас цікавить, ми отримуємо наступний результат:

#### **332.1 Banks**

[332.1 Banking: general resources](#)

[332.1 Central banks](#)

[332.1 Commercial banks](#)

[332.1 International banks](#)

В кінцевому результаті пошуку на згаданому сайті отримуємо перелік тематичних посилань, з одного боку, та анатоційний матеріал щодо цих посилань, з іншого:

332.1 Central banks	
Titles	Descriptions
1. <a href="#">Bank of Canada</a>	<b>Bank of Canada</b> Information in English or French on the Bank of Canada and its publications, exchange rates, library, and currency museum. Author: Bank of Canada DeweyClass: 332.1
2. <a href="#">Bank of England</a>	<b>Bank of England</b> Official information from the central bank of the United Kingdom. The Bank has many roles including the setting of interest rates in order to meet the Government's inflation target, the supervision of banks and wholesale money market institutions, the collection of monetary and banking statistics, management of the Government's stock register, printing of the country's bank notes, and maintenance of close links with the financial markets. Author: Bank of England DeweyClass: 332.1
3. <a href="#">Bank of Japan</a>	<b>Bank of Japan</b> Information about the Japanese economy and the activities of the Bank of Japan. Includes public statements from the Bank, financial statements, economic and financial reports and statistics, research papers and lists of Bank of Japan publications with subscription information. Author: Bank of Japan DeweyClass: 332.1
4. <a href="#">European Central Bank</a>	<b>European Central Bank</b> Provides access to interest rate information, periodical publications, press releases, statistics, key speeches, MFIs (Monetary Financial Institutions), and links to individual EU central banks. Available in a wide range of languages. Author: European Central Bank DeweyClass: 332.1

На нашу думку, даному інтернет-каталозі здійснювати пошук можливо також за глосарієм – небагатовисловним зібранням термінів, з наступним їх функціональним тлумаченням шляхом характеристизації тематичними масивами даного каталогу. Терміни в глосарії зібрано за англійською абеткою.

Пошукові механізми даного сайту дають змогу проводити пошук також і за країнами. Для цього виділено окрему закладку *Countries* – певний перелік країн світу. Сто дев'яносто сім країн складають цей список. Цікавим є представлення країни за певними ознаками вже всередині рубрики, яку подає та чи інша країна. Так, можна побачити:

Russia
<a href="#">arctic</a>
<a href="#">russia</a>
<a href="#">russian business</a>
<a href="#">russian composers</a>
<a href="#">russian history</a>
<a href="#">russian language</a>
<a href="#">russian literature</a>

Italy
<a href="#">italian composers</a>
<a href="#">italian culture</a>
<a href="#">italian history</a>
<a href="#">italian language</a>
<a href="#">italian literature</a>
<a href="#">italy</a>

Тобто відбувається певна характеристизація країн у відповідності до перелічених ознак, що символізує фасетність окремих розділів даного каталогу, побудованого згідно з класифікацією Дьюї.

На сайті, що розглядається, представлена систематизація інфомасивів за типами ресурсів (*Types*):

### Internet resources by type

<a href="#">Bibliographies</a>	<a href="#">Magazine listings</a>
<a href="#">Biographies</a>	<a href="#">Mailing lists</a>
<a href="#">Book and text collections</a>	<a href="#">Map collections</a>
<a href="#">Bookshops</a>	<a href="#">Moving images</a>
<a href="#">Dictionaries</a>	<a href="#">Museums</a>
<a href="#">Directories</a>	<a href="#">Poems</a>
<a href="#">Discographies</a>	<a href="#">Proceedings</a>
<a href="#">Ebooks: fiction</a>	<a href="#">Satellite images</a>
<a href="#">Ebooks: non-fiction</a>	<a href="#">Software archives</a>
<a href="#">Essays</a>	<a href="#">Sound collections</a>
<a href="#">Governments</a>	<a href="#">Thesauri</a>
<a href="#">Image collections</a>	<a href="#">UK higher education colleges</a>
<a href="#">Journal collections and listings</a>	<a href="#">UK universities</a>
<a href="#">Library catalogues</a>	

Що ж до основного ряду, згідно з яким побудовано даний каталог, то він представлений трьохзначними індексами, і має наступний вигляд:

<b>000 Generalities</b> <i>Includes: computing, Internet, libraries, information science</i>	<b>500 Science and mathematics</b> <i>Includes: physics, chemistry, earth sciences, biology, zoology</i>
<b>100 Philosophy and psychology</b> <i>Includes: ethics, paranormal phenomena</i>	<b>600 Technology</b> <i>Includes: medicine, engineering, agriculture, management</i>
<b>200 Religion</b> <i>Includes: bibles, religions of the world</i>	<b>700 The arts</b> <i>Includes: art, planning, architecture, music, sport</i>
<b>300 Social sciences</b> <i>Includes: sociology, politics, economics, law, education</i>	<b>800 Literature and rhetoric</b> <i>Includes: literature of specific languages</i>
<b>400 Language</b> <i>Includes: linguistics, language learning, specific languages</i>	<b>900 Geography and history</b> <i>Includes: travel, genealogy, archaeology</i>

Подальша деталізація рубрик проводиться за класичною структурою класифікації Дьюї: від меншого до більшого.

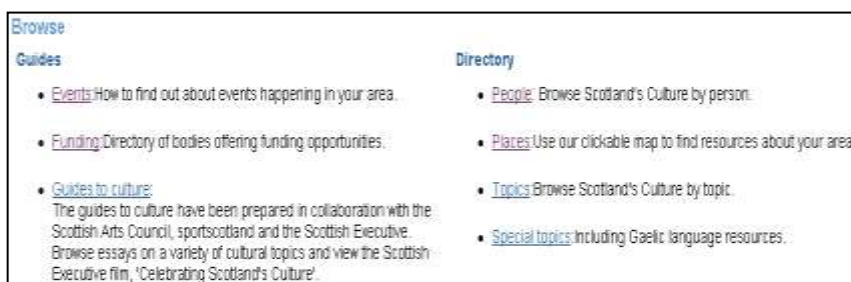
Отже, можна говорити про цілковиту адаптованість класифікаційної ПМ в мережі інтернет. Судячи з вищенаведеної інформації видно, що формування електронного каталогу та пошук в ньому засобами класифікаційної ПМ цілком дієві і широко використовуються на практиці.

В інтернет-середовищі також представлено каталоги, що систематизовано за допомогою предметної ПМ, зокрема *Предметного покажчика Бібліотеки Конгресу США (LCSH)*. Одним з них є ресурс

<http://www.scotlandsculture.org>. [5] – зібрання інформації щодо шотландської культури.

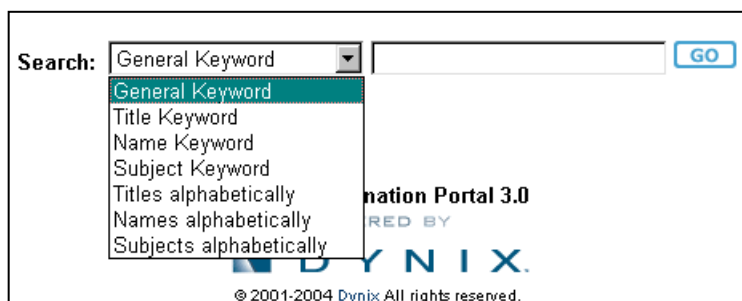
В контексті даного розгляду розділи, що нас цікавлять, це – *Browse* та безпосередньо *Search*.

Перший з запропонованих розділів сайту становить собою перелік ресурсів, що зібрано у відповідності з певними згрупованими узагальнюючими ознаками. Це своєрідна оглядова мапа інфомасиву контекстного характеру. Візуально це представлено так:



З наведеного видно, що інформацію розмежовано залежно від цільового призначення. За це відповідають два підрозділи оглядового представлення *Guides* та *Directory*. Таким чином, використовуючи даний розділ, можна здійснювати пошук інформації не за ключовим словом, автором чи назвою, а безпосередньо за предметною приналежністю інфо-продукту.

За термінологічний пошук в чистому вигляді відповідає другий з наведених розділів. Пошук може відбуватись у відповідності до декількох форм. Так маємо:





Як бачимо, простий пошук відбувається або за основним ключовим словом, назвою, ім'ям, предметом; або за назвою, ім'ям, предметом в алфавітному порядку.

Внаслідок пошукової команди наприклад "music", отримуємо результат у 318 позицій. Цей масив назв-посилань можна впорядковувати, як видно з наведеного прикладу, за назвою, ім'ям, датою публікування, автором.

З запропонованого переліку цілком можливо переглянути короткі



відомості будь-якого електронного ресурсу: автора та організацію, яку він представляє, адресу в інтернеті, анотаційний матеріал тощо.

Далі здійснюється перехід до електронного ресурсу безпосередньо.

Так само як і в фолксонемічному варіанті каталогу, що представлено вище, в розглядуваному ресурсі теж можлива функція відбирання потрібних інфо-документів та формування власного переліку потрібних документів. Це здійснюється за допомогою кнопки "Add to my list", в результаті чого обраний документ відкладається у окремий підрозділ. В результаті такого відбору потрібних джерел формується відповідний список.

Таким чином, видно адекватну застосовуваність предметної ІПМ на практиці і представлення її в веб-орієнтованій системі.

У масштабному інтернет-зібранні інформаційних ресурсів з освіти та науки *Intute* [4] систематизація каталогу проводиться за тезаурусом з мистецтва та архітектури (Art and Architecture Thesaurus), тобто засобами дескрипторної ІПМ. Загальний сервіс *Intute* вміщує в собі чотири окремих



інтернет-сервіси: *Science and Technology, Arts and Humanities, Social Sciences, Health and Life Sciences.*

В кожному з цих сервісів, здійснюється пошук як у відповідності до предметного представлення інформації, так і у відповідності до термінологічного представлення останньої.

Так, можна спостерігати розділ *Browse* в якому викладено предметні категорії. В процесі пошуку відповідно до потрібної предметної категорії відбувається перехід до пошуку за терміном безпосередньо та нижчерівневої предметної категорії.

Щодо термінологічного пошуку то в розділі *Search* представлено можливість широкого пошуку:

Intute > Science, Engineering and Technology > Advanced search

**Subject links**

- About us
- A-Z of services

**Internet catalogue**

- Search
- Browse
- New resources
- Suggest a site

**Internet training**

- Virtual Training Suite
- Support materials
- Intute events

**Additional Services**

- E-journals search
- Hot topics
- Natural hazards
- Science data
- Spotlight
- Subject packs
- Timelines
- World guide

**Advanced search**

Use the form below to construct advanced queries of the Intute database; you can narrow your search by subject, resource type and other fields.

Full details of the [various ways to search Intute Science, Engineering and Technology](#) are available.

If your search of the main Intute database doesn't provide you with enough results, you may wish to try the [Intute Harvester](#). This is a much larger database than the core catalogue, but is of lesser quality

Search: [All fields] for [ ] [Go]

Subjects:

- All subjects
- Astronomy
- Chemistry
- Computing
- Earth sciences
- Engineering
- Environment
- General sciences
- Geography
- Mathematics
- Physics

Resource types:

- All resource types
- Associations
- Bibliographic databases
- Moving images
- News
- Non-bibliographic databases

І на останок, розкриваючи всі основні дієві механізми роботи з сервісом *Intute*, принагідно слід зазначити про можливість перегляду останніх надходжень на сайт. Нові ресурси систематизовано у відповідності до предметних категорій: *Arts and Humanities, Health and Life Sciences, Science Engineering and Technology, Social Sciences.*

Підсумовуючи все викладене вище, компетентно можна говорити про повномірне представлення всіх розглянутих в даній публікації видів ІІМ в

веб-орієнтованих системах. Більше того, аналізуючи наведені приклади, можна говорити про наявність рис одразу декількох ІПМ щодо пошуку інформації на запропонованих сервісах. Відносно ж фундаментальності структурної побудови каталогів, що характеризують ІПМ – належність розглянутих сайтів, то тут слід говорити про чітку відповідність закладеним принципам формування структури впорядкованих масивів.

Слід також зазначити, що спостерігається разюча протилежність між принципами формування каталогів в першому прикладі та трьох інших. В першому прикладі закономірно простежується функція демократичного формування і наповнення каталожного масиву координатними одиницями та тлумачними інформаційними блоками щодо них. В інших – повномірне контролювання формування каталогу від початку до кінця. Таким чином, можна говорити про багатоаспектність щодо принципів формування та використання електронних каталогів створених засобами різних ІПМ, представлених у веб-орієнтованих системах.

#### Використані джерела:

1. Інформація та документація. Інформаційні мови. Терміни та визначення : ДСТУ 2398–94. – К, 1994. – 73 с.
2. BUBL LINK: Catalogue of Internet Resources [Electronic resource]. – Internet resources – Access mode: <http://bubl.ac.uk>. – Title from the screen. – Date of access: 22 september 2007.
3. CiteULike [Electronic resource]. – Internet resources – Access mode: <http://www.citeulike.org>. – Title from the screen. – Date of access: 22 september 2007.
4. Intute: Free online service [Electronic resource]. – Internet resources – Access mode: <http://www.intute.ac.uk>. – Title from the screen. – Date of access: 22 september 2007.

5. SCOTLAND'S CULTURE [Electronic resource]. – Internet resources –  
Access mode: <http://www.scotlandsculture.org>. – Title from the screen. – Date  
of access: 22 september 2007.