

Віра Карнаушенко

## Структура та основні характеристики тезауруса бази даних «Plasma»

*Лінгвістичне забезпечення БД «Plasma» ґрунтується на методі пізнавального, лінгвістичного й комунікативного підходу до розпізнання спеціальної лексики наукової термінології. Дослідження показало, що наукова термінологія з фізики плазми характеризується наявністю великої кількості номінацій-синонімів зі слідами конотативної забарвленості та має свої обмеження на лексичному, синтаксичному й семантичному рівнях.*

*Аналіз спеціальної лексики підмови «фізика плазми» став основою для побудови дескрипторного словника, який містить перелік дескрипторів їх аскрипторів і вказує на парадигматичні відносини між дескрипторами.*

*Тезаурус до БД «Plasma» регулярно поповнюється новими дескрипторами у процесі запровадження даних до цієї БД. Відкрита структура тезауруса дає змогу швидко реагувати на нову інформацію та відбивати зміни, що мають місце в предметній галузі.*

**С**проможність спеціалізованих БД на персональних ЕОМ оперативно реагувати на зміни потреб інформаційного ринку зумовила утворення локальної реляційної БД з фізики плазми для забезпечення інформаційних потреб (ІП) наукових співробітників даної галузі. Необхідною умовою забезпечення доступу до БД завдяки формалізованій природній мові є розробка дескрипторного словника даної предметної галузі.

Інформаційно-пошуковий тезаурус до БД «Plasma» - це контрольований словник лексичних одиниць дескрипторної мови, який базується на лексиці англійської формалізованої природної мови, відображає семантичні відносини між лексичними одиницями і призначений для організації пошуку інформації шляхом індексування документів та запитів.

Тематичний профіль тезауруса БД «Plasma» визначається аналізом ІП користувачів і предметною галуззю бази даних з фізики плазми (БДФП). БДФП розроблена з метою задовольнити ІП фахівців даної галузі науковими матеріалами, надісланими з європейських, американських та японських лабораторій і містить реферати англійських статей за фахом. Збір лексики та формування словника ключових слів відбувається одночасно з поповненням БД новими даними. Словник, тобто вихідний масив термінів підмови «фізика плазми» формується шляхом вибірки ключових слів з рефератів у процесі їх індексації та вводу до БД. Чистота вибірки ключових слів визначає якість наповнення тезауруса дескрипторами та надалі позначається на ефективності пошуку і видачі релевантних запитові документів. Відсутність довідкової літератури й таблиць класифікації, які могли б допомогти у роботі з термінологією даної вузькоспеціалізованої наукової галузі, призвели до того, що головна відповідальність за вірний вибір

ключових слів лежить на адміністраторі БД «Plasma». Тому велике значення на цьому етапі роботи приділяється володінню адміністратором лексикою підмови предметної галузі та регулярним консультаціям з науковими співробітниками. Відбір ключових слів, а також визначення родових і видових дескрипторів тезауруса стали можливими завдяки проведенню лексикологічного аналізу спеціальної української та англійської термінології з фізики плазми і знайомству з етапами термінологізації, запропонованими А.С.Гердом<sup>1</sup>. Сюди належать:

- проведення повної систематизації всіх назв, включаючи всі типи вживання термінів у текстах, у розмовній мові, а також усі синоніми - стандартні, жаргонні, професійно-діалектні;

- розробка чіткої логіко-понятійної моделі терміносистеми, на основі якої оцінюється та уніфікується реально існуюча термінологія. Підсумком цієї роботи є визначення для кожного поняття одного-двох основних термінів, що можуть розглядатися як «кандидати» в стандартні терміни;

- безпосередньо стандартизація термінології. Аналіз логіко-граматичних організацій, дериваційних спроможностей, системності та інших важливих характеристик «кандидатів» у стандартні терміни дасть змогу обрати термін і рекомендувати його до офіційного вживання у словниках, енциклопедіях, керівних документах, офіційних виданнях тощо;

- контроль за процесом термінотворення, встановлення кордонів допустимості застосування іноземних термінів у вітчизняній науково-технічній термінології, вироблення правил аббревіації термінів та ін.

Традиційно предметна галузь знання описувалася фразами, що склалися з іменників і не включали дієслова та інші частини мови. Ця методика заснована

на тому, що в деяких західних мовах (наприклад, англійській) дієслова можна легко перетворити в іменники, використовуючи герундій або віддієслівний іменник. Прагнення фахівців відбити динаміку явища, описаного в документі, призвело до використання дескрипторів інших частин мови, зокрема дієслів (у формі інфінітива).

Аналіз спеціальної лексики наукової термінології з фізики плазми підтвердив наше припущення щодо необхідності відбирати до дескрипторів не тільки окремі номінації-іменники, а й термінологічні словосполучення, які в смисловому відношенні являють собою цільні лексичні одиниці (наприклад, *sawtooth discharge* - розряд з пилкоподібними коливаннями, *giga-landau drift resonance* - гіродрейфовий резонанс Ландау, *particle confinement* - утримання часток та ін.). Незважаючи на те, що в ролі ключових слів БД «Plasma» використовуються іменники та прикметники, в разі необхідності передбачається включення інших частин мови (дієприкметників, прислівників, дієслів). Іменники подаються в називному відмінку однини. Щоб уникнути інформаційних втрат, було розроблено спеціальну проміжну мову даних, що передбачає використання як критерію запиту спеціальної форми дескрипторів (основи слів) та символів-замінників мови SQL.

У процесі дескрипторизації вихідний масив термінів піддається семантичній обробці. Аналіз спеціальної лексики підмови «фізика плазми» стає основою для створення дескрипторного словника, який містить перелік дескрипторів та їх синонімів і вказує на парадигматичні відносини між дескрипторами.

Особливість термінів полягає в їх відношенні до спеціальної понятійної галузі. Понятійна галузь знання описується за допомогою природничих підмов, що мають обмеження на лексичному, синтаксичному й семантичному рівнях. Кожна підмова характеризується своїм набором понять, який організовується термінологом у систему на базі родових, причинних та інших відносин між поняттями. Тому терміни можуть бути реалізовані за допомогою спеціальної лексики в зв'язу концептуальну термінологічну систему.

Хоча терміни характеризують спеціальне поняття певної галузі знання і належать до спеціальної лексики наукової термінології, інколи їх буває важко відрізнити від слів звичайної мови через такі явища, як синонімія, омонімія й полісемія. Крім того, одне й те ж слово може характеризувати різноманітні поняття в різних підмовах та мати відмінне лексичне значення в звичайній розмовній лексиці.

Розпізнанню спеціальної лексики наукової термінології та організації словника дескрипторів даної предметної галузі значно сприяла методика, запропонована Італійською національною дослідницькою радою<sup>3</sup>. Для поліпшення якості пошуку інформації в Інтеграційній юридичній експертній системі розроблено спеціальний тезаурус юридичних термінів Thes Guig. Тезаурус розглядається не тільки як дійова підтримка в роботі інформаційної системи, йому відводиться центральне місце в процесі придбання додаткової інформації, необхідної для використання юридичної експертної системи. На стадії розробки тезауруса було ретельно проаналізовано лексику юридичної термінології. Чіткі критерії щодо відбору дескрипторів дають змогу відмежувати юридич-

ну термінологію від лексики загальноживаної мови і таким чином обмежити ймовірність помилкових зв'язків між дескрипторами та зменшити інформаційний шум у роботі системи. При встановленні лексичних зв'язків між дескрипторами утворені від одного й того ж морфологічного кореня терміни, які не здатні змінити синтаксис фрази та зберігають її смислове значення, визначались як еквіваленти. Наприклад: *adotare* - *приспосовувати*, *adottato* - *приспосований*, *adottante* - *приспосувальний пристрій*, *adottabilita* - *приспосовування*. Враховувалася багато-значність омографічних термінів. Щодо вияву семантичних зв'язків особлива обережність виявлялася при відборі синонімів. Синонімічні відношення зазначалися тільки у випадках абсолютної синонімії. Ще одна група відносин, характерних для тезауруса юридичних термінів - це відношення термінологічної ієрархії зі вказівкою на ширші та вузчі значення термінів.

Визначення парадигматичних зв'язків у тезаурусі БД «Plasma» базується на вивченні лексичних, семантичних та ієрархічних відносин між дескрипторами. Під дескриптором, тобто лексичною одиницею інформаційно-пошукового тезауруса, прийнято розуміти слово чи словосполучення, обране з безлічі умовно-еквівалентних ключових слів для його позначення<sup>2</sup>. Терміни, що мають той самий морфологічний корінь та зберігають смислове значення фрази, визначались як еквіваленти. Наприклад: *discharge phase* - *discharging phase* (*фаза розряду*); *magnetic surface* - *surface of magnet* (*магнітна поверхня*). Еквіваленти, а також дублети, одержані внаслідок абсолютної синонімії, займають у тезаурусі місце дескрипторів. До цієї ж групи віднесено повне найменування загальноживаних аббревіатур, що містяться в тексті реферату.

Пошук даних у БД «Plasma» виконується за тими ключовими словами, які знаходяться безпосередньо в тексті реферату, минаючи процес їх зведення до єдиної форми чи заміни на стандартну лексику даної предметної галузі. Характерність тезауруса до БД «Plasma» в тому, що при віднесенні ключового слова (терміна або термінологічного словосполучення) до групи дескрипторів ми керувалися не значущістю його як сталої лексичної одиниці в термінологічній системі фізики плазми, а смисловим навантаженням, яке несе дана номінація в тексті реферату. Тому до групи дескрипторів було частково віднесено сталі лексичні одиниці наукової термінології, які при формуванні запиту підлягають заміні на дескриптори, представлені термінами-професіоналізмами зі слідами конотативної забарвленості. Інформація, що міститься в колонці «Аскриптори», базується на даних лінгвістичного аналізу наукової термінології з фізики плазми і допомагає користувачеві краще орієнтуватись у спеціальній лексиці предметної галузі в процесі формування пошукового образу документа. У тезаурусі також передбачені посилання на полісемію деяких термінів та термінологічних словосполучень. Наприклад: *rotational transform 1* - *обертальне перетворення*, *rotational transform 2* (*see, toroidal transform angle* - *кут обертального перетворення*). Якщо аббревіатура не виділяється як дескриптор, вона подається в графі дескрипторів після повного найменування термінологічної номінації.

Фізичне подання тезауруса до БД з фізики плазми полягає в утворенні робочого файлу з поширенням mdb (tesaur.mdb) на базі системи керування БД Microsoft Access. Для зручності виводу на екран та використання в процесі пошуку інформації структуру тезауруса подано в двох таблицях дескрипторів: головній (містить перелік родових дескрипторів - найменувань полів БД «Plasma»), та підпорядкованій (з переліком видових дескрипторів і вказівкою на відносини синонімії, омонімії і полісемії між ними). Родові дескриптори становлять тематичні рубрикатори БД, розміщені за значущістю й широтою охоплення тематики предметною рубрикою. Видові дескриптори подані за алфавітом. Наводимо фрагменти таблиць тезауруса.

Зі встановленням реляційного зв'язку між таблицями досягається цілісна картина відбиття парадигматичних зв'язків між дескрипторами. Для зв'язку таблиць використано зв'язок «один до багатьох».

При цьому кожному запису (родовому дескриптору) головної таблиці відповідає один чи більше записів (видових дескрипторів) підпорядкованої. Однак практика показала, що один видовий дескриптор може бути віднесений до кількох родових, а це суперечить організації даних у структурі тезауруса. Тому парадигматичні зв'язки між дескрипторами встановлювалися на основі індексації документа, тобто віднесення його до тієї чи іншої рубрики (тематичного поля БД). При цьому ключові слова документа, внесені до поля тезауруса як видові дескриптори, підпорядковуються тому ж родовому дескриптору-рубрику, що й документ. У разі наявності в документах з різноманітною тематикою того самого ключового слова видовий дескриптор двічі вноситься до таблиці зі вказівкою різних кодів родового дескриптора.

Тезаурус "Plasma"  
Таблиця родових дескрипторів

Код родових дескрипторів	Родові дескриптори	Код родових дескрипторів	Родові дескриптори.
1	автор	17	current drive
2	назва	18	diverter
3	місце публікації	19	pellet injection
4	рік видання	20	diagnostics
5	країна	21	low-high transition
6	реферат	22	equilibrium
7	theory	23	stability
8	experiment	24	confinement
9	numeral experiment	25	magnetic hydrodynamic characteristics
10	tokamak	26	instability
11	stellarator	27	wall-plasma
12	reactor	28	impurities
13	ion-cyclotron resonance heating	29	boundary plasma
14	electron-cyclotron resonance heating	30	ion heatconductivity
15	low hybrid resonance heating	31	electron heatconductivity
16	neutral injection heating		

Таблиця видових дескрипторів

Дивизор	Видові дескриптори	Аскриптори	Код родових дескрипторів
1	advanced tokamak		10
2	Advanced Toroidal Facility, ATF		11
3	antenna		13
	antenna		14
	antenna		15
4	ASDEX	ASDEX tokamak	10
5	banana-plateau regime		24
6	bootstrap current		22
7	closed surface	continuous surface	24
8	charge-exchange recombination spectroscopy, CXRS		20
9	collisionless fluid model		23
10	comercial tokamak reactor		12
11	Compact Helical System, CHS		11
12	core fuelling		19
13	damping force	viscosity force	7
14	DIII-D discharge	discharge in tokamak DIII-D	7
15	density profile		22
16	density profile		23
17	density profile		24
18	direct radiation recombination inverse, DDR-inverse		7
19	dissipative trapped electron mode		7
20	driven current	drow current	17
21	dynamic discharge phase	time depended discharge phase	10
22	dynainic discharge phase	time depended discharge phase	11
23	edge localized mode, ELM		10
24	edge plasma		27
25	electron-cyclotron current drive experiment, ECCD experiment		17
26	electron-cyclotron emission, EC-emission		20
27	electron-cyclotron heating system, ECH system		14
28	electron-cyclotron radiation, EC-radiation		24
29	electron-cyclotron range of frequency, EC-range of frequency		14
30	electron-cyclotron wave, EC-wave		14
31	energy confinement		24
32	electron-cyclotron resonance driven current		17

Головна таблиця тезауруса має поля зі специфікацією:

Код родового дескриптора	числовий
Родові дескриптори	текстовий
Підпорядкована таблиця має такі поля:	
Лічильник	числовий
Видові дескриптори	текстовий
Аскриптори	текстовий
Код родового дескриптора	числовий

Для створення зв'язку був визначений ключ для обох таблиць. Його було привласнено полю «Код родових дескрипторів» головної таблиці, після чого встановлювався зв'язок з одноименним полем підпорядкованої таблиці. При цьому поле «Код родових дескрипторів» головної таблиці виконує функцію лічильника. Пер-

винний ключ у підпорядкованій таблиці було привласнено полю «Лічильник». Визначати первинний ключ при встановленні зв'язку для підпорядкованої таблиці не обов'язково, проте це значно прискорює роботу Microsoft Access. Крім того, для підпорядкованої таблиці було визначене поле вторинного ключа, найменування, тип даних і розмір якого збігаються з полем первинного ключа головної таблиці. На відміну від первинного ключа, для вторинного припускається наявність однакових значень полів. Для цього при наданні специфікації поля вторинного ключа «Код родових дескрипторів» йому була задана властивість «Так - Збіг допускається» у графі характеристик поля «Індексоване поле». Після цього було створено зв'язок між таблицями.

Гере А.С. Терминология-унификация или стандартизация? Научно-техн. инф-я. - 1978. - Сер.2. - №4. - С.14.  
 2. Додова Н.В., Устименко А.И., Петренко Б.В. Информатика. - Минск: Высшая школа, 1990. - С.47.

3. Camelli A., Socci F. A thesaurus for improving retrieval in an Integrated Legal Expert System. Italian National Research Council, Proceedings of the 3-rd Int. ISKO Conf. 1994, V.4-P.242-251.

МАЙСТРИ ХУДОЖНЬОГО ОФОРМЛЕННЯ КНИГИ

Роберт Лісовський

Роберт Лісовський (1893 - ?) з нарбутівської плеяди. Народився в Запоріжжі. З походження - німець. Початкову освіту здобув у німецькій школі. Навчався в Миргородській мистецько-промисловій та в Київській мистецькій школах, в Українській академії мистецтв.

Безпосередніми наставниками художника були О.Сластюн, О.Мурашко, В.Кричевський, М.Бойчук, Ю.Нарбут. Його творча індивідуальність формувалася в колі майбутніх відомих українських митців: В.Крижанівського, В.Седлера, П.Ковжуна, А.Петрицького та ін. Він розробив обкладинки до багатьох книжок, що випускалися видавництвами «Грунт», «Вернигора», «Сяйво», «Музагет» тощо. Особливі роботи - обкладинки до «Сонячних кларнетів» П.Тичини, до оповідання Б.Лепкого «От так собі».

З 1927 р. митець працює в Академії мистецтв у Берліні, в 1929 - 1945 - в Українській пластичній студії в Празі. В останні роки життя мешкав у Лондоні. Учасник багатьох художніх виставок. На жаль, багата спадщина майстра майже невідома на його Батьківщині, в Україні.

Н.С.

