

Виходячи з цієї теоретичної бази, за допомогою систем інтелектуального аналізу даних (ІАД) [3] можна проводити сканування та-ких індивідуально-психологічних особливостей студентів, як «раці-ональність-іраціональність», тип інформаційного метаболізму (ТІМ) та репрезентативна система, на підставі предикатів, які сту-дент вживає при викладі тексту у вільній формі на етапі вхідного контролю.

Інтелектуальний аналіз даних, чи data mining – це процес вияв-лення значущих кореляцій, зразків та тенденцій у великих обсягах даних.

Перевагою застосування подібних технологій у рамках ДН є по-ривильна простота їхнього використання викладачем, яка проявля-ється в тому, що без спеціальних знань основ штучного інтелекту викладач може знаходити «приховані закономірності» в інформації, яка підходить від студента в текстовій формі.

У цей час на вітчизняному та зарубіжному ринках представлена досить велика кількість програмних продуктів, що можуть бути викорис-тани для цієї мети. Наприклад, автоматизація контент-аналізу тексту може здійснюватися за допомогою програмного продукту *TextAnalyst*. Таким чином, контент-аналіз спрямований на витяг *шматків*, які є виявленими закономірностями предметної галузі (принципи, зв'язки, закони), що дозволяють вирішувати задачі ви-ничного класу.

У *TextAnalyst* витягнуті знання представлено у вигляді *семан-тичної мережі* (мережі понять та відносин між ними) деякої пред-метної галузі.

У загальному випадку термін *семантична* означає *значення*, *пояснює семантика* – це наука, яка встановлює відносини між сим-волами й об'єктами, які вони позначають.

Характерною рисою семантичних мереж є обов'язкова наявність *примітивів* відносин [1]: клас – елемент класу; властивість – зна-чення; приклад елемента класу.

Як кількість типів відносин семантичні мережі можуть бути од-норідними (з єдиним типом відносин) та *неоднорідними* (з різни-ми типами відносин).

УДК 378.1

Павло Стефаненко

## ЗАСТОСУВАННЯ ТЕХНОЛОГІЙ «DATA MINING» У ПРОЦЕСІ ДИСТАНЦІЙНОГО НАВЧАННЯ У ВИЩІЙ ШКОЛІ

У процесі дистанційного навчання (ДН) особливу увагу варто приділити скануванню індивідуально-психологічних особливостей студентів, оскільки вони є значущим чинником процесу сприйняття й обробки навчальної інформації. Крім того, акцентування уваги на цьому питанні зумовлене тим, що особистісні характеристики сту-дентів визначають набір індивідуальних критеріїв, до яких до-цільно адаптувати зміст навчальної інформації при розробці на-вчального курсу викладачем.

Зупинимося на критеріях індивідуалізації, заснованих на засто-суванні соціонічного аналізу [1] та НЛП (нейролінгвістичного про-грамування) [1] як елементу, що забезпечує нову теоретичну базу для організації індивідуально-диференційованого підходу до на-вчання.

© Павло Стефаненко, 2001

34

(Семантичні мережі диференціюють також за типами відносин).  
Виходячи з цього критерію класифікації, їх розділяють на два класи:

— **бінарні семантичні мережі**, засновані на відносинах, що зв'язують тільки два об'єкти (поняття тексту);

— **парні семантичні мережі**, які містять спеціальні відносини, що зв'язують більше двох понять.

Загальна класифікація семантичних мереж зображена на рис. 1:

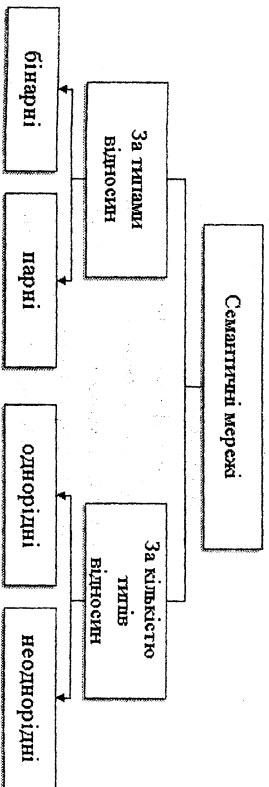


Рис. 1. Класифікація семантичних мереж

Поряд з тим зазначимо, що відносини семантичних мереж реалізують визначені типи зв'язків. Ті з них, які використовуються найбільш часто, представлені на рис. 2:

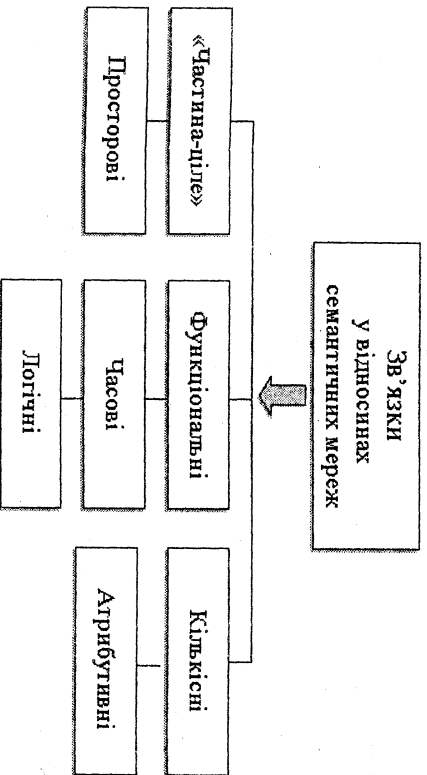


Рис. 2. Зв'язки у відносинах семантичних мереж

Зв'язки типу «частина-ціле» подібні до зв'язків «клас-підклас», «елемент-безліччя» і т. ін.

**Функціональні зв'язки** в тексті традиційно визначаються дієсловами («визначає», «впливає», «сприяє» тощо).

Для визначення кількісних зв'язків між поняттями тексту використовуються порівняльні предикати (більше, менше, дорівнює). Аналогічно визначаються інші типи зв'язків: **просторові** (далеко від, близько від, за, під, над), **тимчасові** (раніше, пізніше, протилежно), **атрибутивні** (мати властивість, мати значення), **логічні зв'язки** (і, чи, не).

Перейдемо до безпосереднього розгляду функціональних можливостей програми TextAnalyst, що можуть бути використані викладачем у процесі сканування індивідуально-психологічних особливостей студентів, які навчаються за дистанційною формою (Рис. 3).

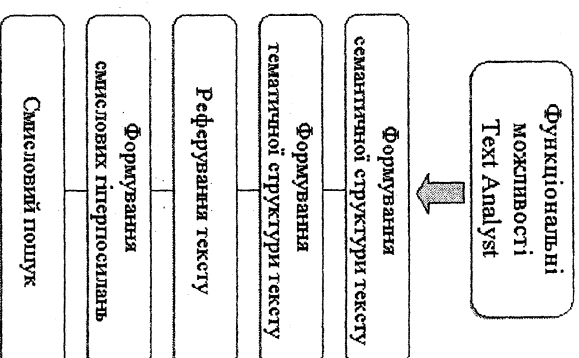


Рис. 3. Функціональні можливості TextAnalyst

Алгоритм роботи цієї програми представлений сукупністю етапів:

1. На підставі запропонованого для аналізу тексту **програма формує передусім семантичну мережу**, до складу якої включаються

поняття, які несуть основнеислове навантаження. Крім значущих понять, у мережі відображаються найбільш інформативні смислові зв'язки між поняттями. Таким чином, на цьому етапі мережа дозволяє відкинути несуттєву інформацію та представити зміст у «стиснутому» вигляді.

Результатом цього етапу алгоритму є надання можливості виділити всю інформацію, пов'язану з кожним поняттям — темою тексту за допомогою смислових зв'язків.

2. До кожного елементу семантичної мережі (поняття), а також до кожного зв'язка між парами понять ставиться у відповідності, **смислова вага** у вигляді певної числової оцінки. Ці оцінки дозволяють:

- оцінити частку «внеску» поняття чи зв'язку в семантику тексту;
- з'ясувати ступінь деталізації опрацьованої в тексті тематики;
- задати спосіб сортування інформації;
- досліджувати текст за визначеними рівнями чи «паррами» — смисловими зрізами різної глибини.

У результаті цього етапу алгоритму програма формує «дерево понять», у якому кожному з них привласнюються розглянуті ваги. **Значення смислової ваги** (від 1 до 100) показує, наскільки важлива роль цього поняття в структурі змісту всього тексту (іншими словами, кількість інформації, що стосується цього поняття). Максимальне значення ваги (100) поняття свідчить про те, що воно є в тексті **ключовим**.

Нагомість мале, близьке до 1, значення показує, що тема, яка пов'язана з цим поняттям, лише поверхово згадана в тексті. Тобто з цією темою пов'язано дуже мало інформації, а ступінь інформативності поняття є низьким.

Діапазон зміни ваги зв'язку в програмі також має розмірність [1...100]. Велике значення ваги зв'язку між вкладеними поняттями в **дереві понять** вказує на те, що велика частина інформації в тексті, яка пов'язана з першим поняттям, стосується й другого. Це означає, що **розглянута тема майже завжди викладається в контексті іншої**.

Мале значення цього коефіцієнта свідчить про обмежений виклад теми в контексті обох понять. Користувач (викладач) може довільно

визначувати способи сортування під час обробки тексту, **імнісаного** студентом.

Як ілюстрацію наведемо семантичну мережу, сформовану **програмою TextAnalist** для окремого фрагмента тексту, в якому описується еволюція дидактичних систем (рис. 4).

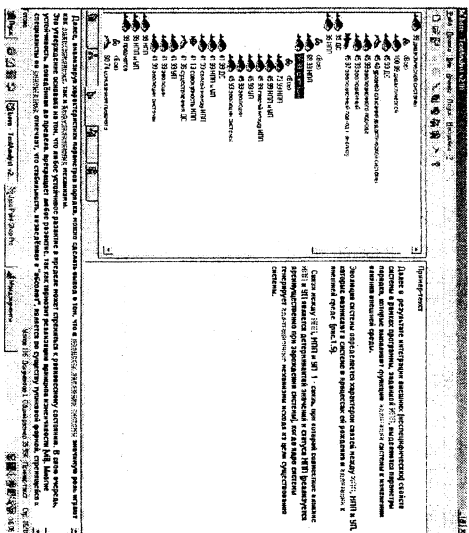


Рис. 4. Семантична мережа фрагмента тексту

З ілюстрації видно, що основнеислове навантаження цього фрагмента тексту несуть такі поняття, як «дидактична система», «еволюційний підхід», специфічні параметри порядку (НППД), керуючі параметри (КП), зв'язки між інваріантними параметрами порядку (ППД). Їхня смислова вага (значення, розміщене ліворуч від поняття) дорівнює 99. Крім того, з рис. 4 видно, що у контексті поняття «адаптація» переважно трапляється термін НППД (вага всіх зв'язків дорівнює 73), і майже у два рази рідше — «зв'язки між НППД», КП, еволюція системи (ваги зв'язків рівні 45).

Крім семантичної структури тексту, програмою формується **тематична** структура тексту. Цей режим є найбільш зручним для сприйняття користувачем, тому що представляє текст у найбільш структурованому вигляді з виділенням головних і другорядних **попів**.

Ієматична структура — це представлення змісту тексту у вигляді ієрархії понять. При цьому зв'язки між поняттями є **односпрямо-ваними** (від головного поняття до доручорядного). Таким чином, тематична структура має вигляд дерева, в корені якого розташовані головні поняття, а **галузі** являють собою підлеглі підтеми.

Програма надає можливість регулювати **ступінь зв'язності** тематичного дерева. Зміна порога за вагою зв'язків у мережі понять (розірвання більш-менш сильних зв'язків) змінює вид дерева, розбиваючи його на визначене число **тематичних вузлів**.

Поряд із викладеними вище функціями викладач може за допомогою програми **TextAnalist** виконувати функцію **реферування** написаного студентом тексту чи швидкого ознайомлення з його змістом (рис. 5). У результаті програма за текстом, що аналізується, формує **реферат**, який містить список найбільш інформативних речень текстів — тез, розташованих за частотністю вживання в тексті.

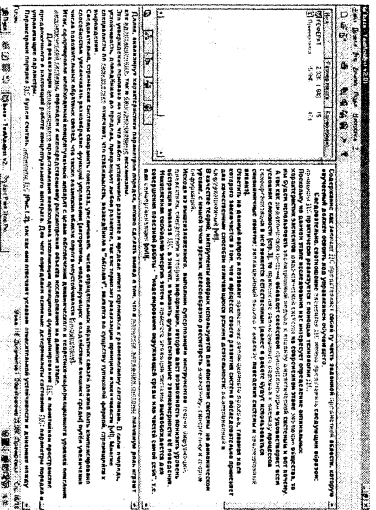


Рис. 5. Реферат аналізованого фрагмента тексту (праве верхнє вікно)

Усі речення реферату мають посилення на вихідний текст. Таким чином, з'являється можливість переглянути контекст потрібної тези. Ступінь деталізації реферату може налаштовуватися шляхом регулювання кількості формулювань його речень, причому кожному реченню також надаються ваги.

Працюючи з програмою **TextAnalist**, користувач може самостійно формувати **гіпертекст**. На відміну від традиційних гіпертексто-

вих систем, у яких усі маршрути руху по тексті апріорно «нав'язуються» розробниками систем, програма TextAnalist дозволяє автоматично перетворити об'ємний масив текстової інформації в **гіпертекст із можливими гіперпосиланнями**, виділивши всі потенційно присутні взаємозв'язки та переходи.

Алгоритм формування гіпертексту охоплює два кроки: формування мережі понять, проєкцію мережі понять на вихідний текст.

У результаті цього на основі обраного фрагмента тексту з'являється можливість циклічного данцюгового руху (рис. 6).

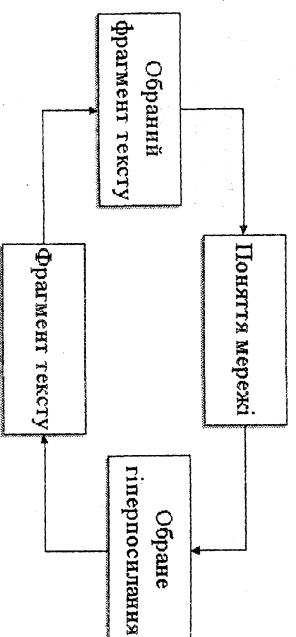


Рис. 6. Формування гіперпосилань за допомогою смислових взаємозв'язків

І, напевні, найскладнішою функцією програми є **«смисловий пошук»**. Вона дозволяє одержати відповідь на запит, сформуований у вигляді фрази природної мови або словосполучень із набору ключових слів. Причому інформація, що «виглядається», може мати не тільки іншу граматичну форму, але й взагалі не згадуватися в тексті шуканому, однак зберігати з ним смисловий зв'язок.

Для ілюстрації функції програми «смисловий пошук» звернемося до розглянутого вище прикладу. Для обраного фрагмента тексту ніслемо запит **«рівні опису дидактичної системи»**.

Результатом смислового пошуку за цим запитом є основна інформація, що стосується рівнів опису дидактичної системи та представлена в правому верхньому вікні програми (рис. 7). У сформованому текстовому фрагменті виділені поняття, які несуть основне смислове навантаження.

П. Стефаненко

**Применение технологий «data mining» в процессе дистанционного обучения в высшей школе**  
Резюме

В статье рассмотрены возможности применения технологии интеллектуального анализа данных (data mining) как фактора повышения уровня интерактивности в процессе обучения в высшей школе на примере программной продукта TechAnalyst. Особое внимание уделено функциональным возможностям программы, а именно: формированию семантической и тематической структуры текста, реферированию текста и смысловому поиску.

P. Stefanenko

**On Using «Data Mining» Technology in Distance Learning at Higher School**  
Summary

The article deals with the ways of applying data intellectual analysis technologies (data mining) as a factor of increasing interactivity level during distance learning in a higher school on the example of TechAnalyst software product. Special attention is paid to functionalities of the program, namely: to the text semantic and subject pattern forming, to the text abstracting, and to semantic search.

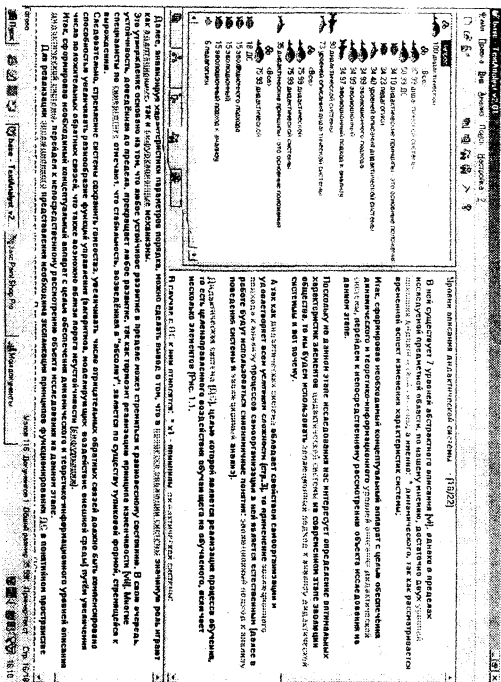


Рис. 7. Результат обработки запиту програмою TechAnalyst

Отже, у результаті аналізу функціональних можливостей програми TechAnalyst можна зробити висновок про те, що вона як елемент штучного інтелекту в рамках дистанційного навчання може використовуватися для таких цілей:

— семантичної обробки результатів тестів третього та четвертого рівня засвоєння для оцінки знань студентів;

— семантичної обробки результатів вхідного контролю, представлених у вигляді текстових файлів, із метою виявлення предикатів, на підставі яких викладачем можуть бути визначені особистісні характеристики студентів.

Розглянутий варіант автоматизації цих процесів, на наш погляд, є способом підвищення ступеня інтерактивності й ефективності дистанційного навчання.

1. Молодцов А., Хохель С. Практикум по прикладной социологии. — К.: МАУП, 1994. — 204 с.
2. Додер Х. НЛП: Современные психотехнологии. — СПб: Издательство «Питер», 2000. — 160 с.
3. Демонстрационная версия программы TechAnalyst\*.
4. Информационные технологии: Материалы Сибирского государственного университета путей сообщения\*.

\* статті розміщені в мережі INTERNET.

Стаття надійшла до редакції 07. 06. 2001