Марк Бургін

Інфологічна система як розвиток поняття «тезаурус»

Так само, як енергія і маса завжди існують у формі матеріальних об'єктів, так і інформація згідно з положеннями семантичної текстової теорії інформації існує у формі якогось тексту (повідомлення). У загальному значенні текст - це сукупність виразів якої-небудь мови. Не всяка знакова система - це мова, але будь-який знак, якщо він зрозумілий, несе інформацію. Тому її містять не тільки тексти, а й інші агрегати зі знаків. Приміром, природа, картини, фотографії. Виникнення текстів пов'язане зі специфікою сприйняття і передачі інформації людиною. Поняття тезаурусу, з одного боку, дуже важливе для вивчення інформації, а з другого,- недостатнє для цього, бо інформація - це джерело не тільки знань, а й ідей, уявлень, образів тощо. Тому подальший розвиток теорії інформації потребує розширення та узагальнення поняття «тезаурус». Таким узагальненням стало поняття «інфологічна система».

Структурні інфологічні елементи допускають визначення як ідеальні семантичні (або символьно-знакові) системи різних рівнів: від окремого символу або знака і до цілої мови. З іншого боку, будь-яка система структур може трактуватись як структурна інфологічна система.

Термін «тезаурус» грецького походження і означає «скарбниця». Зараз у це поняття вкладають значення «вся кількість накопичених знань». За прийнятою термінологією, в кібернетиці тезаурус означає сукупність знань окремої людини чи певного колективу [1]. У вужчому смислі - це будь-який словник, призначений для пошуку слів якоїсь мови за їх змістом, що в явному вигляді фіксує семантичні відношення між одиницями, з яких він складається [2]. А найрозповсюдженіше розуміння тезаурусу таке: це словник слів та виразів, згрупованих згідно зі спільними аспектами їхніх значень [3].

Тезаурус компактно представляє мовний простір та системні відношення в ньому, дає змогу спостерігати рівневу організацію лексики, є наближеною інтерпретацією лексичної системи, її формальною моделлю, і дозволяє зобразити системні відношення, спостерігати рівневу організацію лексики, тобто є способом відображення семантичного простору.

З іншого боку, словом *тезаурус* послуговуються для позначення компонентів автоматизованих систем. Так, він є головним інструментом інформаційно-пошукових систем (ІПС). У цьому разі тезаурус - це автоматизований словник, що відбиває семантичні відношення між лексичними одиницями дескрипторної інформаційно-пошукової мови (ІПМ), та призначений для автоматичного пошуку слів за їх змістом. Конкретизація зумовлює виділення різних типів тезаурусів: базового, багатогалузевого, вузькотематичного та ін. [4]. В інформатиці термін «тезаурус» використовується також для близьких за змістом, але все ж різних понять. Наприклад у [5] тезаурус - це словник синонімів.

Крім того, в практиці науково-технічної інформації під цим поняттям, як правило, розуміють формалізований тезаурус великої системи і не завжди чітко розмежовують тезауруси й словники. Тим часом у широкому значенні тезаурус уособлює семантичне наповнення системи, котре нераціонально розглядати поза поняттям «система» [6]. Взагалі, у найширшому значенні в словнику української мови тезаурусу найбільше відповідає одне зі значень слова «світ»: наприклад, *світ підлітка*, *світ робота* [7].

Поняття тезаурусу використовується в семантичній теорії інформації [8]. Її, на відміну від семантичної логічної теорії інформації [9], природно називати текстовою. Для неї вихідними є поняття *тексту* і *тезаурусу*. Текст T є іменованою множиною T=(G,t,A), у котрій G - орієнтований граф, тобто теж іменована множина [10]; A - якийсь алфавіт, тобто множина символів; і t: V(G) - A відображення множини V(G) вершин графа G в A. Набір текстів з одним алфавітом називається знаковою системою. Тезаурус B - це довідник, в якому записана інформація про смислові зв'язки в якійсь мові L, де мова - це клас знакових систем, між котрими визначені співвідношення заданого виду [8].

Так само, як енергія і маса завжди існують у формі матеріальних об'єктів, так і інформація згідно з положеннями семантичної текстової теорії інформації існує у формі якогось тексту (повідомлення) [8]. У загальному значенні текст - це сукупність виразів якої-небудь мови. Не всяка знакова система - це мова, але будь-який знак, якщо він зрозумілий, несе інформацію. Тому її містять не тільки тексти, а й інші агрегати зі знаків. Навіть якщо взяти розширене розуміння тексту із [8], то інформацію містить і природа, і картини, і фотографії. Виникнення текстів пов'язане зі специфікою сприйняття і передачі інформації людиною.

Інформація в семантичній теорії інформації також не визначається, а до теорії вводиться кількість інформації I (T,B) як міра змін у тезаурусі B під дією тексту (повідомлення) T. Крім того, на відміну від статистичної (шеннонівської) теорії інформації, в семантичній текстовій теорії інформації, за [8], ще немає розвиненого математичного апарату, з яким можна ефективно працювати.

© Бургін Марк Семенович, Київ, 1997

Усі конкретні реалізації тезаурусів виходять з того, що все знання може бути виражене в понятійній формі. Водночас сучасні дослідження з методології науки [11,12] демонструють, що поряд з понятійним існують й інші форми знання: декларативне, гіпотетичне, проблемне тощо. Тому загальне визначення тезаурусу як місця зберігання знань вступає у суперечність з тим, що його конкретизації включають тільки поняття та зв'язки між ними.

Справжня «скарбниця» знань має складатися з кількох підсистем понятійного або лінгвістичного, логіко-дедуктивного або, ширше, логіко-аргументаційного, модельного чи репрезентативного, процедурного або операційного, аксіологічно-нормативного тезаурусів. Тільки перший з них відповідає звичайному тезаурусу. Всі наступні включають інші типи знань. Але все це не повністю відображає всі можливості мозку людини.

Як бачимо, з одного боку поняття тезаурусу дуже важливе для вивчення інформації, а з другого - недостатнє для цього, бо інформація - це джерело не тільки знань, а й ідей, уявлень, образів тощо. Тому подальший розвиток теорії інформації потребує розширення та узагальнення поняття «тезаурус». Таким узагальненням стало поняття *інфологічна система*.

Підсистема F системи R називається інфологічною, якщо вона складається з інфологічних елементів: повідомлень, даних, знань, образів, схем, текстів, уявлень, вірувань, абстракцій, ідей тощо. Ці елементи певним чином (за допомогою зв'язків) організовані в єдину систему. Найтиповішим прикладом інфологічної системи є те, що наповнює мозок людини, або сам цей мозок. Пам'ять комп'ютера, ІПС, бібліотека, архів дають інші приклади інфологічних систем.

Згідно з формою існування, можна виділити інфологічні елементи різних типів: структурні, матеріальні, реляційні. Структурні елементи, в свою чергу, бувають іконічними, семіотичними та синтетичними. Іконічні інформаційні елементи є або прямими відображеннями будь-чого засобами інфологічної системи, або породженнями самої цієї системи, які подібні (мають таку ж форму) до прямих відображень. Семіотичні інфологічні елементи поділяються на такі різновиди, як знання, гіпотези, ідеї, правила, норми, уявлення, вірування, семіотичні образи тощо. І нарешті, синтетичні інфологічні елементи складаються як з іконічних, так і з семіотичних.

Структурні інфологічні елементи допускають визначення як ідеальні семантичні (або символьно-знакові) системи різних рівнів: від окремого символу або знака і до цілої мови. З іншого боку, будь-яка система структур може трактуватись як структурна інфологічна система.

У свою чергу, поняття структури теж має різні трактування. У теорії систем структура виступає як певна складова системи. А саме, система складається з елементів та зв'язків між ними. Ось ці зв'язки і називаються структурою даної системи. У теорії тріад структура визначається як інформаційний варіант систем, об'єктів, процесів. У інфологічних системах структури мають символьно-знакову форму.

Щодо знаків, то необхідно зробити три уточнення. По-перше, як відомо [13], слово «знак» використовується в науковій літературі різнопланово: це показник, мітка або те, що може виражати, подавати ідею. Ним може слугувати предмет, людина, явище, процес.

Це і «матеріально» зображена заміна предметів, явищ, понять у процесі обміну інформацією в людській групі, оскільки основна суть знака - здібність реалізовувати функцію заміщення. Усі такі знаки є матеріальними. Але є й ідеальні знаки. Наприклад, слово «знак», написане тричі, є трьома матеріальними знаками, але тільки одним ідеальним. Ідеальні знаки формально відповідають класам еквівалентних матеріальних знаків. Не обов'язково, що кожен такий клас задано конструктивно. Тому вважається, що ідеальний знак існує, коли існує хоча б один його матеріальний представник. Саме це мається на увазі, коли пишеться літера «а» чи слово «або».

Кожен знак є певною системою. Він може бути або не може бути підсистемою якоїсь більшої системи. Знакові або семіотичні системи теж бувають матеріальними (наприклад, текст, який ви читаєте) або ідеальними (наприклад, будь-яка мова).

По-друге, часто вважають, що знак - це обов'язково елемент певної системи [13]. Але така умова занадто обмежує, бо природніше виділяти системні та позасистемні (типу символів) знаки [14]. У такому разі будь-який образ чого-небудь теж виступає в ролі знака. Але природно розділяти іконічні знаки, подібні до того, що вони відображають, та умовні знаки, які поділяються на символи та семіотичні (системні) знаки. Поготів, одне й те саме може виступати в ролі різних знаків (навіть знаків різних типів), в залежності від того, яке значення пов'язується з цим об'єктом.

По-третє, існує можливість, що якісь об'єкти утворюють систему, побудовану за правилами семіотики, але її елементи нічого не позначають. Прикладом є формальні мови, які широко вивчають у математичній лінгвістиці. Елементи таких семіотичних систем називатимемо потенційними або формальними знаками, на відміну від актуальних знаків, які щось позначають.

Взагалі існує також ситуативний підхід до визначення знака. У цьому випадку будь-який об'єкт є знаком тільки тоді, коли він перебуває у знаковій ситуації, яка потребує наявності [15]:

1) предмета, що за певних умов виконує функцію знака;

2) предмета, до якого він відсилає;

3) смислового значення (сліду), за допомогою якого реалізується відсилка;

4) організованої системи, що відсилається до певного предмета.

Таким чином, структурні інфологічні елементи - це ідеальні сутності (структури, форми), які належать до структурного рівня світу [16].

Матеріальні інфологічні елементи є носіями структурних інфологічних елементів. Приміром: листок паперу з якимись позначками; книга; магнітофонна або фотоплівка, якщо на ній щось записано або сфотографовано; елемент пам'яті комп'ютера.

Але матеріальні об'єкти виступають носіями структурних інфологічних елементів завдяки тому, що вони перебувають у певному стані або в них здійснюється зміна станів. Так, елементи пам'яті сучасних комп'ютерів - тригери - перебувають у двох станах, один з яких

відповідає знаку *1*, другий - знаку *0*. Коливання частинок повітря (тобто зміна їх станів) несе звукову інформацію. Відзначимо, що повітря, з одного боку, є каналом зв'язку, а з іншого - місцем зберігання (хоч і на дуже короткий час) структурних інфологічних елементів. Такі ж дві функції виконує й екран дисплея чи телевізора. Саме стани або їх зміна і є реляційними інфологічними елементами.

Поняття інфологічної системи є базовим для загальної теорії інформації [17,18]. Згідно з цією теорією, є інформація в широкому смислі та у вузькому. А саме, за другим онтологічним принципом, інформація (у вузькому смислі) I для системи R - це будь-яка сутність, котра спричинює зміни в інфологічній системі F системи R.

Саме в цьому виявляється, насамперед, дійовий характер інформації. Але тільки на базовому рівні світу - структурному [16]. Наскільки це важливо, демонструє те, що люди часто змішують поняття даних, знань, інформації, що спричинює багато непорозумінь. Новий підхід (тобто загальна теорія інформації) дає змогу прояснити ситуацію. Дані і знання - це структури, котрі містять інформацію так само, як і матеріальні об'єкти - енергію. Саме тому одні й ті ж знання (чи дані) порізному сприймаються різними людьми. Все залежить від того, яку інформацію видобудуть люди зі знань. Важливо не просто дати людині знання, а навчити її вилучити з них правильну інформацію.

Це ж стосується і повідомлень, стосовно яких прямо пишуть, що вони саме і є інформація. Насправді ж повідомлення тільки можуть містити інформацію. Телеграфний апарат, котрий друкує телеграму, приймає не інформацію (в строгому смислі), а тільки повідомлення. Отримати інформацію, що міститься в цьому повідомленні, здатна лише система, що має інфологічну підсистему: людина або комп'ютер, але не всякий, а тільки той, який має відповідне програмне забезпечення. Тому вірним є твердження про те, що комп'ютери обробляють інформацію. Навіть зміною форми даних можна отримати нову інформацію.

Згідно з головними принципами математична модель окремої порції інформації будується у вигляді оператора, заданого на математичних структурах, котрі відображають розглядувані системи. Для побудови математичної моделі фіксують можливі або реальні класи систем-носіїв і систем-приймачів інформації. З онтологічних принципів випливає, що оператори подання інформації в математичній моделі визначаються на структурах виду (C,in,R), де C - система-носій інформації (або її математична модель), R - система-приймач інформації; in - взаємодія між ними. Такі структури називаються іменованими множинами або тріадами [10,16]. Вони використовуються в математичній частині загальної теорії інформації.

Отже, будь-яка бібліотека, база даних або знань є інфологічною системою. Якщо базу знань (а тим більше, базу даних) цілком припустимо розглядати як певний тезаурус (але тільки в розширеному розумінні), то довільна бібліотека в цілому виходить поза його межі, даючи приклад більш загальної інфологічної системи. Справді, зміст більшості художніх творів (романів, поем, віршів тощо) не зводиться суто до знань, а є перважно системою уявлень, образів, міркувань. Теж саме стосується багатьох повнотекстових або гіпертекстових систем, не кажучи вже про мультимедіа, комп'ютерні системи. Тим самим інфологічна система дає значно більшу й точнішу модель бібліотеки або гіпертекстової системи, ніж тезаурус.

1. Словарь по основам информационной деятельности. - К., 1995.
2. *Hornbiy A.S.* Oxford Advanced Learners Dictionary of Current English. - Oxford, 1980.
3. *Карaулов Ю.Н.* Лингвистическое конструирование и тезаурус литературного языка. - М., 1981.
4. *Морковкин В.В.* Идеографические словари. - М., 1970.
5. *Сэлтон Дж.* Эксперименты по автоматическому построению тезаурусов // Киб. сб. - 1974. - №11. - С. 185-199.
6. *Воробьев Г.Г.* Теория тезаурусов в анализе коммуникаций // Семиотика и информатика. - М., 1979. - Вып. 11. - С. 52-65.
7. *Воробьев Г.Г.* Теория тезаурусов в документалистике // Вопросы кибернетики: формализация языковых данных. - М., 1977. - С. 10-19.
8. *Шрейдер Ю.А.* О семантических аспектах теории информации // Информация и кибернетика. - М.: Сов. радио, 1967. - С. 15-47.
9. *Bar-Hillel Y., Carnap R.* Semantic Information // Brit. J. Philsc. 1958. - V. 4 - №3. - P. 147-157.
10. *Burgin M.S.* Theory of Named Sets as a Foundational Base for Mathematics // Structures in Mathematical Theories. - San Sebastian, 1990. - P. 417-420.
11. *Бургин М.С., Кузнецов В.И.* Аксиологические аспекты научных теорий. - К.: Наук. думка, 1991.
12. *Бургин М.С.* Феномен знания // Филос. и социолог. мысль. - 1995. - №1/2.
13. *Салмина Н.Г.* Знак и символ в обучении. - М.: Изд-во Моск. ун-та, 1988. - С. 288.
14. *Бургин М.С., Кузнецов В.И.* Мир теорий и могущества разума. - К., 1992.
15. *Троицкий В.П.* Семиотическое существование // Семиотика и информатика - М., 1979. - Вып. 11. - С. 3-36.
16. *Бургін М.С.* На шляху до абсолюту // Вісн. НАН України. - 1993. - №5.
17. *Burgin M.S.* Dynamic Theory of Information // Abstr. of the AMS. - 1991. - V. 12. - №1.- P.199.
18. *Бургин М.С.* Алгоритмический подход в динамической теории информации // Докл. РАН. - 1995. - Т. 342. - №1. - С. 7-10.