

Список використаних джерел:

1. Інструкція з розрахунку наявної пропускної спроможності залізниць України ЦД-0036 : наказ Міністерства транспорту та зв'язку України від 23.11.2002 р. № 1026. – К. : Мін-во трансп. та зв'язку України, 2002. – 376 с.
2. Кочнев Ф. П. Организация движения на железнодорожном транспорте / Кочнев Ф. П., Акулиничев В. М., Макарович А. М. – М. : Транспорт, 1979. – 568 с.
3. Тимчасова інструкція з організації швидкісного руху пасажирських поїздів. Вимоги до інфраструктури та рухомого складу : наказ УЗ від 12.07.2002 № 360-Ц.
4. Бабаєв М. М. Вплив швидкості руху поїздів на вимоги до інформаційного забезпечення локомотивних систем сигнального авторегулювання і пропускної здатності залізничних ліній / М. М. Бабаєв, С. В. Кошевий, А. С. Зайцев // Збірник наукових праць УкрДАЗТ. – 2010. – Вип. 118. – С. 12–21.
5. Лисенков В. М. Статистическая теория безопасности движения поездов : учеб. для вузов / Лисенков В. М. – М. : ВИНТИ РАН, 1999. – 332 с.
6. Вернигора Р. В. Аналітичний розрахунок коефіцієнтів зйому вантажних поїздів пасажирськими в умовах швидкісного руху / Р. В. Вернигора, О. Ю. Папахов, Н. О. Логвінова // Східно-Європейський журнал передових технологій. – 2013. – № 2/3 (62). – С. 51–55.
7. Козаченко Д. М. Моделювання роботи залізничного напрямку / Д. М. Козаченко, Г. Я. Мозолевич, О. В. Власюк // Вісник Дніпропетровського національного університету залізничного транспорту імені академіка В. Лазаряна. – 2009. – Вип. 28. – С. 144–148.
8. Козаченко Д. М. Дослідження параметрів потоків поїздів на залізничних напрямках / Д. М. Козаченко, Г. Я. Мозолевич // Восточно-Европейский журнал передовых технологий. – 2010. – № 5 (45). – С. 17–21.
9. Інструкція зі складання графіка руху поїздів на залізницях України: ЦД-0040 [Електронний ресурс] : наказ УЗ від 05.04.2002 № 170-Ц. – К. : Транспорт України, 2002.



УДК 004.4'414

Б. І. Мороз, доктор технічних наук, декан факультету інформаційних та транспортних систем і технологій Академії митної служби України

Д. Є. Костенко, аспірант кафедри інформаційних систем та технологій Академії митної служби України, провідний фахівець лабораторії

інформаційних систем та процесів в митній справі

В. В. Костенко, викладач кафедри інформаційних систем та технологій Академії митної служби України

**ВИКОРИСТАННЯ ОНТОЛОГІЙ В УПРАВЛІННІ ЗНАННЯМИ
ДЛЯ ПОКРАЩАННЯ ІНФОРМАЦІЙНОЇ СФЕРИ У СУСПІЛЬСТВІ**

Розглядаються проблеми та можливості управління знаннями на основі масивів неструктурованих текстів у службових документах. Аналізується можливість використання онтологій для розв'язання проблем пошуку інформації та службових документів. Досліджується питання старіння інформації.

Ключові слова: база даних/знань; база документів; дані; документ; інформаційна система; інформаційна сукупність; інформація; онтології; пошукові системи; пошук документів; семантичний пошук.

© Б. І. Мороз, Д. Є. Костенко, В. В. Костенко, 2014

There are examine the problems and management possibilities by knowledge on the basis of the unstructured texts arrays in official documents. There are affected the question about possibility of the use of ontologies as a decision of problems for search information and official documents. There are examine the question of information aging.

Key words: database/knowledge base; base documents; data; document; informative system; information set; information; ontologies; searching systems; search documents; semantic search.

Постановка проблеми. Сучасний етап розвитку суспільства супроводжується швидким зростанням інформації та інформаційних потоків, які безпосередньо впливають на розвиток та життєдіяльність будь-якого підприємства або організації.

Виникають нові типи відносин між людьми, нові концепції щодо організації соціуму, зокрема інформаційне суспільство, що є соціологічною концепцією, котра визначає головним фактором розвитку суспільства виробництво і використання науково-технічної та іншої інформації. Концепція інформаційного суспільства є різновидом теорії постіндустріального суспільства. Завдяки можливості спілкування (за участю великих масивів інформації), в інформаційному суспільстві класи змінюються “інформаційними спільнотами”.

Аналіз останніх досліджень і публікацій. Найсучасніший математичний апарат, який дозволяє описувати ситуації, відношення та приймати складні рішення в умовах невизначеності в нечітких множинах запропонував американський математик Л. А. Заде [1].

В основі підходу лежить ідея про те, що системи, які залежать від людського сприймання, та реакції нечіткі за своєю природою, тому їх аналіз має базуватись на понятті нечіткої множини. Нечіткі множини, які оперують лінгвістичними змінними, дають можливість описувати “людські знання” та оперувати ними. У цій теорії мірою невизначеності для оцінки інформації служить характеристика, яка ґрунтується на погляді про “ступінь належності”, який виказується в нечіткій формі думки і називається в теорії нечітких множин лінгвістичною змінною. Значенням такої змінної служать слова або речення мови. Застосування лінгвістичних змінних сприяє можливості приблизної оцінки складних явищ, для яких немає чіткого переходу від одного стану до іншого [1].

З розвитком новітніх технологій усе більш помітним став їхній вплив на людину, суспільство в цілому.

Мета статті – аналіз, дослідження та покращання можливостей ефективного пошуку інформації в сукупності неструктурованих документів; дослідження перспектив створення спеціалізованої програмної системи з використанням спеціальних методів і алгоритмів морфологічного й синтаксичного аналізу неструктурованих текстів у службових документах, оптимізації пошуку інформації та її статистичної обробки. Перевіряються питання, які пов’язані зі старінням інформації. Важливе також питання використання онтологій (у тому числі для розв’язання проблем старіння інформації). Вважаємо доцільним використання принципів семантичного пошуку в базах даних/знань для забезпечення точного результату.

Виклад основного матеріалу. Більша частина часу користувача витрачається на пошук інформації. Однак ефективність “розкопок” в інформаційній сукупності різна: одна людина знаходить інформацію миттєво, в іншій на це витрачається дуже багато часу, ну а третя може й зовсім не знайти для себе нічого корисного. У чому ж причина? Можна дати досить просту відповідь: потрібно знати, де шукати.

Наскільки точні результати будуть отримані насамперед залежить від того, наскільки вміло сформований пошуковий запит. Наприклад, якщо ведеться пошук документа, пов’язаного з технічною інформацією, більше цінних відомостей можна знайти, якщо зробити пошук за ключовими словами. У цьому випадку користувач передбачає, що в документі такі слова обов’язково знайдуться.

Недолік полягає у тому, що в різних напрямках технічної сфери застосовуються ті ж самі поняття, які користувач часто використовує як ключові. Крім того, процес пошуку може відбуватися в досить великій базі даних. Буває, що навіть на конкретно сформований запит (саме користувачами), може бути досить неточний результат пошуку. Це залежить від структурованості або неструктурованості даних. Неструктуровані дані становлять більшу частину інформації, з якою мають справу користувачі. Ці дані становлять не менше 90 % усієї інформації, а 10 % – це структуровані, впорядковані дані.

Знайти в цій плутанині корисні матеріали можна лише за допомогою спеціалізованих технологій. Основу розвитку останніх становлять дослідження й досягнення у моделюванні (математичному, інформаційному тощо). Інтегруючись у комп'ютерні технології, моделі (досліджуваного об'єкта, предметної області або процесу) набувають якостей і властивостей досліджуваної системи, що допомагає вивчати особливості об'єкта, а також розробляти їх зміни. З іншого боку, комп'ютерні технології усе ширше стають істотно важливим інструментом для моделювання.

Кожному зрозуміло, що навіть у величезних масивах інформації мають працювати пошукові системи, які забезпечували б користувачеві результат – швидкий і точний. Сучасне інформаційне суспільство використовує ряд аналітичних підходів до подання інформації, щоб забезпечити її подальший пошук [2]. Деякі базуються на теорії множин, інші – на елементах векторної алгебри. Обидва підходи ефективно реалізуються в умовах практики.

Пошук є однією із центральних процедур у системах автоматичної обробки інформації. Виконуючи будь-яке завдання, потрібно вибирати вихідні дані з певних заздалегідь підготовлених масивів інформації та записувати результати в ці або інші масиви. Найраціональнішим підходом до організації масивів інформації під час виконання обчислювальних завдань є концепція банків і баз даних. Проблеми пошуку інформації у формалізованих базах даних заслуговують найсерйознішої уваги.

Пошук інформації в базах даних – це відбирання з них множини описів об'єктів, що задовольняють сформульовані в запиті умови. При цьому як результати пошуку можуть видаватися не всі ознаки об'єктів, а тільки частина їх – відповідно до умов запиту. Об'єкт може вибиратися з масиву за значенням однієї ідентифікаційної ознаки (ключової) або за поєднання значень декількох ключових ознак. Об'єкт може також вибиратися за поєднання будь-яких інших (неключових) ознак, якщо воно однозначно виділяє його з множини всіх об'єктів масиву.

Важлива методологічна проблема прикладної комп'ютерної лінгвістики полягає у правильному оцінюванні необхідного співвідношення між декларативною і процедурною компонентами систем автоматичної обробки текстової інформації. Чому віддати перевагу: потужним обчислювальним процедурам, що спираються на відносно невеликі словникові системи з багатою граматичною й семантичною інформацією, чи потужному декларативному компоненту з досить простими процедурними засобами?

Інформація достовірною, якщо вона показує реальний стан справ. Об'єктивна інформація завжди достовірною, але достовірною інформація може бути як об'єктивною, так і суб'єктивною. Достовірною інформація допомагає прийняти нам правильне рішення.

На жаль, достовірною інформація, як і будь-яка інша, піддається процесу старіння. Процес старіння додає купу проблем під час пошуку необхідних документів. Старінню інформації полягає в зменшенні її цінності із часом. Старінню інформації забезпечує не сам час, а поява нової інформації, яка уточнює, доповнює чи відкидає повністю або частково більш ранню. Науково-технічна інформація старіє швидше, естетична (твори мистецтва) – повільніше. Із часом кількість інформації зростає, інформація накопичується, відбувається

її систематизація, оцінка й узагальнення. На жаль, обсяги зростають дуже швидко. Нині це терабайти текстових даних.

Старіння інформації в різні моменти часу формально можна зобразити у вигляді:

$$S_t(t) = M_t - M_{t'},$$

де $S_t(t)$ – старіння інформації (даних) зареєстрованих у момент часу t' на момент t ; $M_t, M_{t'}$ – значення даних відносно досягнення мети в момент часу t, t' .

Моделі пошуку визначають ступінь відповідності між запитом і результатом. Взагалі можна вивести загальну ідею: більший відсоток збігу між запитом і документом дає зрозуміти, що документ більше відповідає запитує користувача. Поняття “збіг” багатоаспектне: включається термінологічна база, ключові поняття, збіги предметних областей.

Поняття онтології, запозичене з філософії, зараз активно застосовується в штучному інтелекті й інформатиці. У філософії онтологія вивчає категорії буття, які існують або можуть існувати. У штучному інтелекті онтології згадуються в контексті з такими поняттями як концептуалізація, знання, подання знань, що ґрунтуються на знаннях системи.

Одні намагаються дати неформальні визначення, інші описують онтології на основі понять і конструкцій логіки й математики. Хоча побудовано безліч різних онтологій і збільшується область їхнього застосування, дотепер немає точного визначення цього поняття стосовно області штучного інтелекту [3].

Будь-яка база знань, система, що ґрунтується на знаннях, фіксується явно або неявно деякою концептуалізацією. Множина об'єктів і відношення між ними відбиваються в словнику, в якому система, що базується на знаннях, показує свої знання. Таким чином, вважається, що основу онтології становлять множини її термінів. Утім, не тільки термінів. Як уже говорилося, в онтологічну сукупність включаються також відомості про предметні області, про області визначень тощо.

Х. Такеда ставить онтології в центр проблеми організації знань, тому що в кожній області можуть існувати різні розуміння тих самих термінів. У цьому випадку онтологія використовується для структурування інформації, залишаючись посередником між людиномашинно-орієнтованим і машинномашинноорієнтованим рівнем подання інформації. Тоді онтологія визначається як якась “угода” про деяку область інтересів для досягнення певних цілей [4].

На думку Н. Гуаріано, для встановлення взаєморозуміння про знання, які представлені на деякій мові, зокрема логічній, онтологія має характеризувати концептуалізацію, обмежуючи можливі значення предикатів і функцій. У цьому розумінні, онтологія – це логічна теорія, аксіоми якої обмежують інтерпретації нелогічних символів мови [5].

Неформально онтологія являє собою деякий опис погляду на світ стосовно конкретної сфери інтересів. Цей опис складається з термінів і правил використання цих термінів, що обмежують їхні значення в рамках конкретної галузі. На формальному ж рівні онтологія – це система, що складається з набору понять і набору тверджень про ці поняття, на основі яких можна будувати класи, об'єкти, відносини, функції й теорії. Онтологія як приклад загальної угоди про семантику області сприяє встановленню коректних зв'язків між значеннями елементів області, тим самим створюючи умови для їх спільного використання.

Онтології можна застосовувати як:

- 1) будівельні блоки компонентів баз знань;
- 2) окремі блоки впливу;
- 3) схеми об'єктів в об'єктно-орієнтованих системах;
- 4) структурований глосарій взаємодіючих блоків системи;
- 5) словник для зв'язку між елементами;
- 6) спосіб визначення класів для програмних систем.

Процес побудування онтологій може містити такі етапи [3]:

1. Фіксація знань стосовно предметної області:

- визначення основних понять та їх взаємовідносин в обраній предметній області;
- створення точних визначень (які не конфліктують між собою) для понять і відношень;
- визначення термінів, які пов'язані між собою відповідними відношеннями;
- остаточне погодження всіх вищеназваних етапів.

2. Кодування:

- процес поділу сукупності окремих термінів (які використовуються в онтології) на окремі класи понять;

- вибір або розробка спеціальної мови для представлення онтології;

- створення фіксованої концептуалізації на обраній мові представлення даних.

Онтології пропонується застосувати як посередника між користувачем і процесом пошуку, між процесом пошуку і пошуковою системою. Будову онтологічної системи можна зобразити за допомогою діаграм (рис. 1, 2).

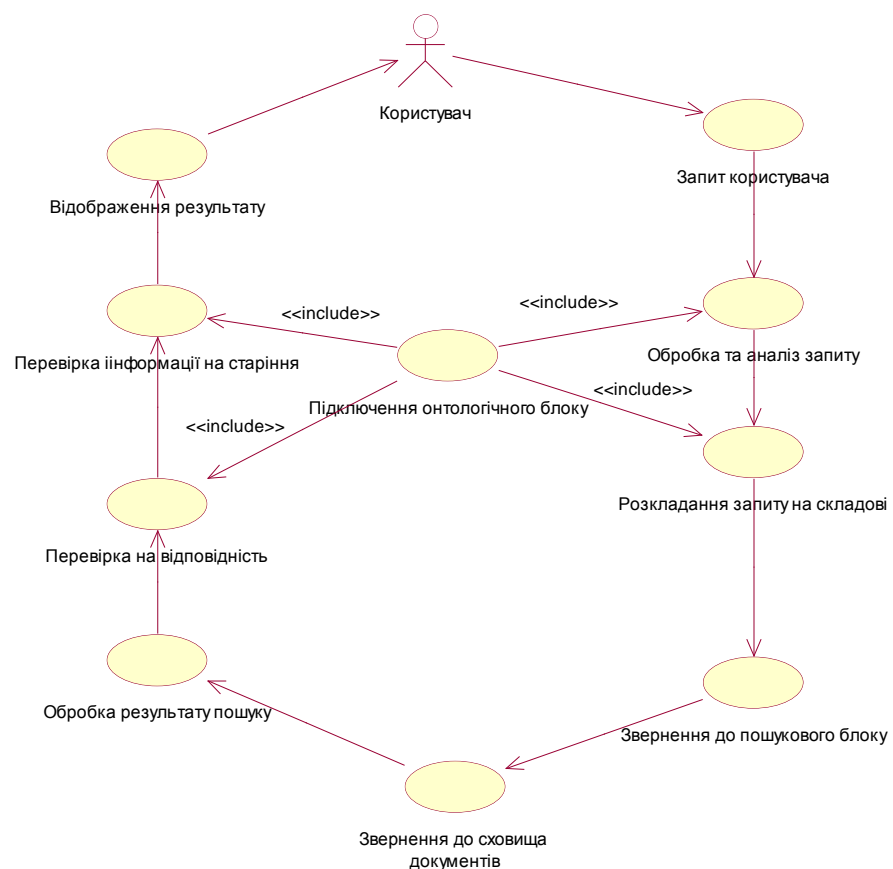


Рис. 1. Діаграма варіантів використання системи онтологічного пошуку

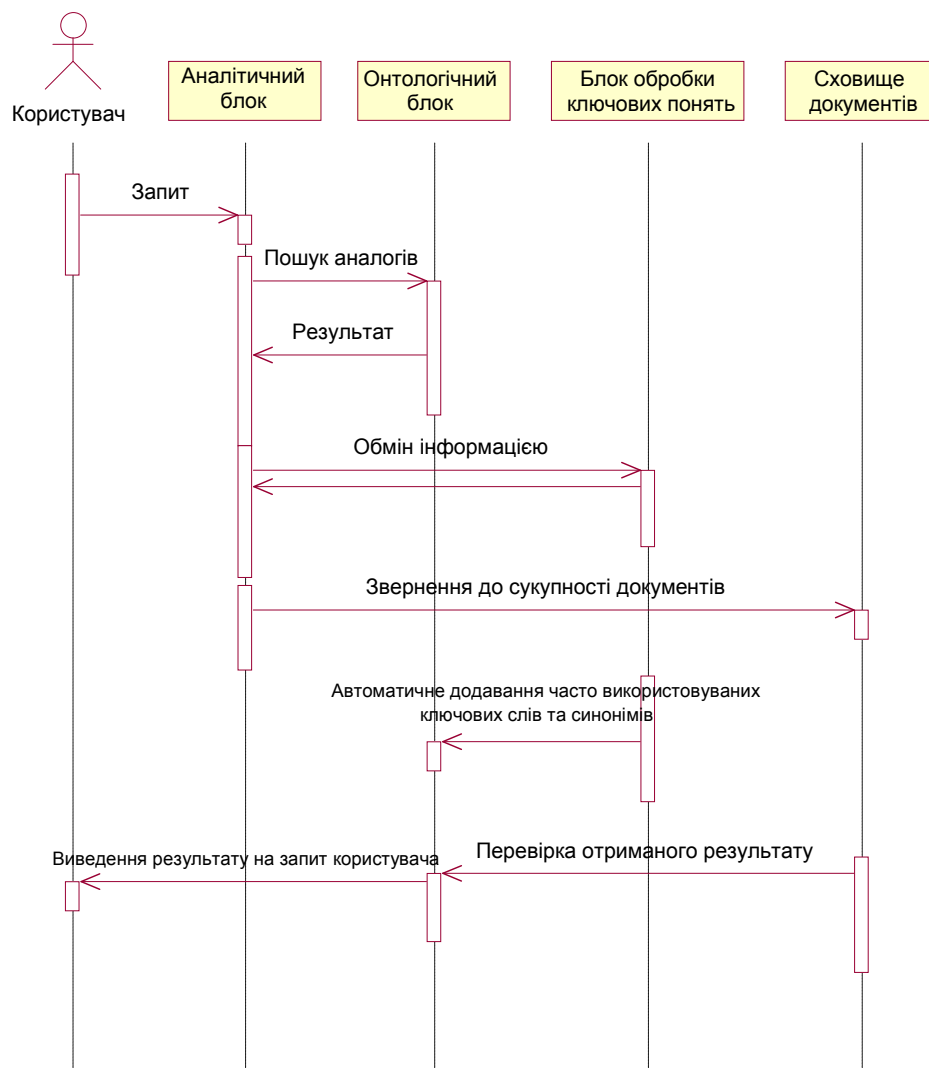


Рис. 2. Діаграма послідовності для варіанта використання “Запит”

Звертаючись до пошукової системи, користувач повинен мати можливість одержати у відповідь ресурси, які релевантні змістові запиту, а отже, пошук має бути семантично орієнтованим. Тому засоби пошуку інформації пропонується організувати на основі онтології, що містить опис семантики ресурсів.

Для побудови онтології потрібне формальне декларативне подання чітко організованих конструкцій, які містять словник термінів тематичної області, опис визначень цих термінів, наявні взаємозв'язки між ними і взагалі теоретично можливі й неможливі взаємозв'язки.

Взаємодія з онтологією вбачається на таких етапах:

- 1) обробка та аналіз запиту;
- 2) розкладання запиту на складові;
- 3) перевірка інформації на старіння та відповідність;
- 4) перевірка на старіння інформації, яка була обрана зі сховища даних.

Проблема полягає в тому, щоб зробити пошук динамічним і зручним для користувача. Для будь-якого типу запиту, що виникає в практичній діяльності, мають бути знайдені адекватні знання в інформаційному просторі. При цьому мова для формулювання пошукової вимоги не повинна бути занадто складною.

Зокрема, спілкування користувача з пошуковою системою можна зробити простішим, наблизивши мову запиту до природньої.

За такої організації пошуку на етапі формування образу релевантного документа з користувачького запиту виділяються значеннєві структури: значущі слова й терміни предметної області. Ці значеннєві структури потім використовуються для формування окремого пошукового образу під впливом онтологічної складової. На етапі побудови запиту до пошукової системи здійснюється виведення на онтології. При цьому виконується перетворення користувачького запиту в поєднаний логічними зв'язками набір термінів і понять, котрі використовуватимуться пошуковою системою.

Будь-яким складним системам (зокрема, системам пошуку інформації) притаманні фундаментальні властивості, що потребують застосування системного підходу в їх дослідженні (як варіант – дослідженні методами моделювання). Передбачається, що будь-яка модель даних повинна мати подібні властивості.

Це такі властивості:

1) цілісність (система розглядається як єдине ціле, що складається із взаємодіючих елементів, можливо, неоднорідних, але одночасно сумісних між собою);

2) зв'язність (наявність істотних стійких зв'язків між елементами та їх властивостями, причому із системних позицій значення мають не будь-які, а лише істотні зв'язки, що визначають інтегративні властивості системи);

3) організованість (наявність певної структурної та функціональної організації, що забезпечує зниження ентропії системи порівняно з ентропією системотвірних факторів, які визначають можливість створення системи, до яких можуть належати: кількість елементів системи, кількість істотних зв'язків, котрими може володіти кожний елемент, тощо);

4) інтегративність (наявність якостей, властивих системі в цілому, але не властивих жодному з її елементів окремо, тобто властивості системи хоча й залежать від властивостей елементів, але не визначаються ними повністю).

Висновки з даного дослідження і перспективи подальших розвідок у даному напрямку. Онтології можна назвати однією зі спроб порятунку сховищ інформації від тотального переповнення й виникнення помилок. Передбачається, що онтології: можуть і могли б розв'язувати проблему подання знань для виведення інформації, що релевантна запитові користувача, дозволили б виконувати фільтрацію та класифікацію інформації, створювати загальну термінологію для програмних агентів і користувачів.

Перевіряється припущення, що онтології могли б розв'язати таку проблему “старіння інформації”.

У перспективі планується розробка спеціалізованої програмної системи, що дасть можливість використовувати спеціальні методи й алгоритми морфологічного і синтаксичного аналізу неструктурованих текстів службових документів, централізовано зберігати отриману інформацію, оптимізувати пошук інформації та статистично обробляти її.

Крім того, планується розробка методу й алгоритму семантичного пошуку з урахуванням особливостей обробки масивів неструктурованих текстів у службових документах. Планується розробка моделі семантичного пошуку.

Список використаних джерел:

1. Заде Л. А. Понятие лингвистической переменной и его применение к принятию приближенных решений / Заде Л. А. – М. : Мир, 1976. – 168 с.
2. Аникин В. М. Аналитические модели детерминированного хаоса / Аникин В. М. – М. : Физматлит, 2007. – 328 с.
3. Россеева О. Организация эффективного поиска на основе онтологий / О. Россеева, Ю. Загорулько // Труды международного семинара Диалог'2001 по компьютерной лингвистике и ее приложениям. – Аксаково, 2001. – Т. 2. – С. 333–342.
4. Takeda H. Collaborative development and Use of Ontologies for Design / H. Takeda, M. Takaai, T. Nishida // Proceedings of the Tenth International IFIP WG 5.2/5.3 Conference PROLAMAT 98. – Trento, Italy. – September 9–12.
5. Guarino N. OntoSeek: Content-Based Access to the Web. / N. Guarino, C. Masolo, G. Vetere // IEEE Intelligent Systems – 1999. – May/June. – P. 70–80.