

УДК 378:004(07)

Стрижак Олександр Євгенійович, радник директора Інституту обдарованої дитини НАПН України, кандидат технічних наук, м. Київ, e-mail: stryzhak953@gmail.com

Дем'яненко Валентина Борисівна, аспірантка Інституту інформаційних технологій і засобів навчання НАПН України, керівник відділення «Комп'ютерні науки» Київської малої академії наук учнівської молоді, м. Київ, e-mail: demyanyenko@i.ua

КОМП'ЮТЕРНІ ОНТОЛОГІЇ – ТЕХНОЛОГІЧНА ОСНОВА ФОРМУВАННЯ ОСВІТЯНСЬКИХ ІНФОРМАЦІЙНИХ РЕСУРСІВ

Анотація

Описуються підходи, моделі і засоби формування систем знань з метою подальшого розвитку освітнянських інформаційних ресурсів. Розглянуто онтологічний аспект проектування баз знань навчального призначення. Визначаються онтологічні моделі, засоби лінгвістичного, структурного і функціонального аналізу й побудови описів об'єктів навчальних дисциплін. Проведення досліджень у вказаному напрямку передбачає вирішення актуальних проблем підвищення ефективності використання освітнянських інформаційних ресурсів на основі застосування сучасних інформаційно-комунікаційних технологій за допомогою засобів дистанційного доступу до розподілених систем знань.

Ключові слова: інформаційна модель, онтологія, морфологія, тезаурус, лінгвістичний і структурний аналіз, база знань.

Постановка проблеми в загальному вигляді. Одним із перспективних напрямків подальшого вдосконалювання електронних систем навчання є розробка методологічних, онтологічних і логічних основ конструювання баз знань навчального призначення (БЗНП) [2, 3, 6, 9, 15]. До онтологічних аспектів відноситься коло питань, починаючи від сфери застосування й до формального опису компонентів комп'ютерних онтологій предметних дисциплін. Головний вектор досліджень спрямований на формалізацію етапів побудови, структурування й подання матеріалу предметної дисципліни під час навчання й інтегрованого з інформаційним навчальним ресурсом проблемного простору, що дозволяє отримати ефективне поєднання засвоєння лекційного й лабораторного матеріалів. У свою чергу,

ефективна реалізація зазначених етапів і одержання кінцевого результату (у вигляді бібліотеки онтологічних баз знань предметних дисциплін) неможлива без проведення системно-онтологічного аналізу заданої сукупності інформаційних навчальних ресурсів.

Аналіз останніх досліджень і публікацій та виклад основного матеріалу.

Сучасні дослідження визначеної проблематики в основному розглядають засоби опрацювання мовних даних. Ці засоби будуються на основі «внутрішньомовної» і «комп'ютерної» сферами опрацювання і виконуються у ході реалізації відповідного алгоритму, що використовує базу «мовних знань» (основою якого є мовно-онтологічна картина світу (МОКС)) і базу знань заданої предметної області (ПдО). Ланцюжок інформаційних технологій – *комп'ютерне опрацювання природномовних текстів* → *представлення знань* → *комп'ютерне опрацювання бази знань* представляє реалізацію базових процедур аналізу, синтезу й розпізнання природної мови комп'ютером, які в більш загальному розумінні можна виразити продукційним ланцюжком *вхідне повідомлення* → *система знань* → *результат*. Суть цього ланцюжка визначається міждисциплінарною системною інтеграцією лінгвістичних і предметних знань, що взагалі, є новою інформаційною технологією, яка знаходиться в стадії становлення й інтенсивного розвитку досліджень [9, 13, 16]. Але педагогічні й методичні аспекти використання цих технологічних рішень взагалі потребують окремого дослідження. Це обумовлено тим, що інформаційні ресурси нині досить інтенсивно розвиваються і це має значний вплив на сучасний освітній процес. Умови формування освітянських інформаційних ресурсів описані в дослідженнях, що викладено в наступних джерелах – [1–3, 4, 5, 8, 9, 10–14].

У цій роботі розглянуто онтологічний аспект проектування баз знань навчального призначення, який є одним із важливих практичних застосувань напрямку онтологічного інжинірингу. За допомогою запропонованого підходу можуть бути вирішені, а також поліпшені результати розв'язання наступних актуальних завдань в галузі освіти:

- автоматизована розробка баз знань навчального призначення на основі лінгво-семантичного аналізу великих обсягів текстової інформації з використанням оригінальних інструментальних засобів. Водночас вихідна текстова інформація може

використовуватися з різноманітних джерел, наприклад з апробованих у навчальних закладах підручників українською і російською мовами з предметної дисципліни;

- структурування термінів і понять, що містяться в інформаційних навчальних ресурсах з визначеної предметної дисципліни;
- істотне зменшення трудомісткості складання баз знань навчального призначення.

Комп'ютерні онтології – це розділ інформатики, як теоретичної, так і практичної, що інтенсивно розвивається [3, 8, 9, 14]. Комп'ютерну онтологію деякої предметної дисципліни можна розглядати як загальнозначущу, відкриту базу знань, що представлена загальноприйнятою (формальною) мовою специфікації знань. В онтолого-класифікаційній схемі засобів і методів штучного інтелекту онтологічний підхід трактується як різновид системного підходу, заснованого на знаннях. Онтологічний підхід забезпечує ефективне проектування компонентів будь-якої знання-орієнтованої інформаційної системи.

Електронний навчальний ресурс повинен мати всі структурні елементи, що визначають аксіоматизацію наукових знань: аксіоми або схеми аксіом; вихідні (первинні) поняття; правила виведення й побудови системи знань; визначення; теореми, наслідки, положення, що виведені з теорії, логіко-методологічні принципи аксіоматизуючої теорії.

Довільна онтологія (проста, змішана або формальна) ефективна, насамперед, у навчанні. Дійсно, набагато ефективніше показати учневі онтограф предметної дисципліни, ніж описувати концепти дисципліни і їхні зв'язки з іншими засобами. Онтологія структурує знання деякої предметної галузі відповідно до онтологічних принципів. Викладачеві онтологія (особливо комп'ютерна) полегшить розробку навчального курсу, студентові – поліпшить сприйняття навчального матеріалу й міжпредметних зв'язків. Виходячи з цього, можна чекати значного ефекту покращення якості формування знань предметної дисципліни шляхом застосування онтологій в освіті. Найбільш повно переваги онтолого-керованої інформаційної системи навчального призначення проявляються під час використання і взаємодії онтологій двох рівнів – онтології домену предметних дисциплін і онтологій самих предметних дисциплін.

Використання онтології як компонента інтерфейсу користувача особливо важливо під час автоматизованого створення баз знань навчального призначення, тому що онтологія включає семантичну інформацію щодо обмежень, накладених на поняття і їх взаємозв'язок. Онтології, наприклад, можуть бути застосовані для генерації інтерфейсів БЗНП, що відповідають за перевірку порушення обмежень. Викладач і учень можуть ознайомитися з онтологією, щоб краще зрозуміти термінологічний словник, використовуваний у предметній дисципліні.

Реалізація зазначених технологій потребує урахування різних формально-методологічних вимог, критеріїв і оцінок. Наведемо основні з них.

1. Побудова інформаційної й функціональної моделей предметних дисциплін.
2. Необхідність структурування термінів і понять.
3. Правила формування достовірних тверджень і висновків, що описують терміни й поняття предметних дисциплін.
4. Підтримка онтології домену предметних дисциплін.

Під час розробки інформаційного навчального середовища одним із перших постає завдання опису понятійної структури відповідної навчальної дисципліни. Важливо, що у процесі навчання чи підготовки дослідження можна вивчати не лише окремий термін (поняття), але й отримувати всі семантичні зв'язки з іншими поняттями, тим самим осмислюючи його роль у даній системі знань чи в ході розв'язання задачі. Цим і обумовлене використання тезауруса – упорядкованої множини основних термінів-понять даної навчальної дисципліни й характерних для неї семантичних зв'язків між поняттями [2, 5, 6, 15].

Однак тезаурус може виступати не лише засобом організації бази знань. Розширюючи його функції, тезаурус можна перетворити на середовище, у якому забезпечується активізація роботи, а також оригінально розв'язуються навчальні завдання. Крім декларативних властивостей знань, які описуються тезаурусною моделлю, її засобами можна також подавати зміст об'єктів навчальної дисципліни через їх конструювання. Причому як для викладача, так і для учня проектується режими роботи на основі відомої схеми «адміністратор – користувач».

Різні проблемні ситуації в ході розв'язування задач у тезаурусі можна моделювати на понятійному рівні – не вказуючи конкретні числові значення,

метричні співвідношення та ін. Не спрощуючи процес розв'язання задачі приведенням його лише до підстановки числових величин у певні параметри, зазначимо, що моделювання задачі на більш узагальнюючому рівні має й більш загальне значення, ніж лише знаходження конкретного результату, а, значить, і ширшу основу для інтерпретації.

Окремим важливим етапом є добір основних понять для навчального тезауруса. Оскільки його розробка одним лише експертом із предметної галузі (наприклад, викладачем) є складним процесом, необхідна тісна співпраця між спеціалістом з управління базами знань й експертом на всіх етапах роботи [2, 15]. Ця взаємодія може відбуватися двома способами. У першому випадку спеціаліст з управління базами знань розробляє перший варіант тезауруса відповідно до певної методики, консультуючись з експертом щодо змісту знань. У другому випадку експерт з предметної галузі сам розробляє тезаурус, користуючись запропонованою методикою і консультуючись зі спеціалістом з управління базами знань щодо правильного її застосування. Добір понять для навчального тезауруса і визначення семантичних зв'язків між ними є доволі складним процесом і проводиться за спеціальною методикою. Побудова тезауруса здійснюється в кілька етапів:

- визначення навчальної дисципліни;
- складання словника (добір понять). На цьому етапі необхідна кропітка робота зі спеціальною літературою, відповідними тлумачними словниками (якщо такі є), підручниками, методичними матеріалами, навчальними посібниками, стандартами тощо;
- визначення понять;
- визначення переліку семантичних зв'язків між поняттями даної навчальної дисципліни;
- тлумачення семантичних зв'язків (визначення їх змісту);
- побудова семантичної мережі зв'язків понять;
- конструювання схеми словникової статті тезауруса (поняття, його коротке визначення, перелік семантичних зв'язків з іншими поняттями);
- формування складу словникових статей тезауруса.

Побудова тезауруса – надзвичайно складний і трудомісткий процес. Кожен етап пов'язаний з аналізом багатьох варіантів. А кінцевий результат роботи є новим

інтелектуальним продуктом. Для цього необхідно забезпечити реалізацію певних умов до яких, у першу чергу, треба віднести таке:

- чітке розуміння мети і напрямку навчальної діяльності – постановка завдання;
- достатня інформаційна база;
- повний опис об'єктів предметної галузі відповідно до технологічної платформи (технологічною платформою можуть бути такі системи: Convera, Exalead, Галактика Zoom, Інформбюро та інші);
- організація роботи експертів і/або аналітичних груп заданої теми предметної галузі.

Окрім цього, необхідно окремо зазначити, що формування знань про об'єкт можливе лише за умови єдності трьох компонентів:

- сукупність достатніх масивів інформації;
- знань і досвіду експертів і фахівців (аналітиків, методистів, викладачів);
- ефективного аналітичного інструментарію (наприклад, Convera, Exalead, Галактика Zoom, Інформбюро тощо).

Під час добору даних і розробки лінгвістичних ресурсів (ЛР) необхідно враховувати, що є об'єктом дослідження – предметна галузь, її тематичний розділ, процеси, властивості, функціональний опис тощо. Об'єкт є особливим пізнаваним предметом, блоком або єством (реальним або абстрактним), що має важливе функціональне призначення в даній предметній галузі. Об'єкт має структуру, властивості, стан, виявляє чітку функціональність, може мати межі. Група чи множина об'єктів, що мають зв'язки і пов'язані спільною структурою і функціональністю можуть бути об'єднані в класи.

Дослідження інформаційного простору з погляду отримання якнайповніших даних про предметну галузь передбачає таке:

а) добір даних про об'єкт в цілому:

- назва об'єкта;
- його структура і клас;
- близьке оточення;
- інформаційні джерела;
- далеке оточення;

- зв'язки і відношення тощо;

б) добір даних про об'єкт в контексті мети, що визначена в навчальному процесі:

- опис об'єктів його тематик як сфер навчальної діяльності;
- дослідження функціональних властивостей;

в) добір даних про об'єкт з погляду його визначення:

- опис сфер його застосування як простору вирішення завдань;
- опис перетину з іншими об'єктами;

г) аналіз і оцінку зібраних даних;

д) складання інформаційного портрета об'єкта (агреговані описи, аналітичні записки, структурний опис, перелік властивостей тощо);

е) подальший моніторинг інформаційного поля з метою постійного поширення інформаційних описів об'єкта.

Створення інформаційної бази даних – ключовий момент зазначеного процесу. Бази даних мають бути представлені в електронному вигляді й відповідати поставленим завданням.

У цілому інформаційна база даних будується:

а) на основі:

- зовнішніх джерел даних;
- внутрішніх джерел даних;

б) за каналами:

• внутрішніми каналами даних (локальні ресурси навчального закладу, друковані видання тощо);

• відкритими каналами даних (доступні в Інтернеті бібліотечні ресурси, ЗМІ (друковані видання, електронні); відеоджерела тощо);

в) на доборі:

• максимально повної бази даних про об'єкти дослідження, виходячи з поставлених завдань й обраних джерел інформації.

Теми онтологічних описів визначаються поставленими завданнями навчального процесу. Для створення якнайповнішої картини дослідження необхідно проробити й передбачити можливі застосування об'єкта і його складових як під час

добору даних, так і розробки лінгвістичних ресурсів, які повинні забезпечити створення бази знань про об'єкт.

Весь процес добору, опрацювання, аналізу інформації і синтезу вже отриманих знань є низкою певних послідовних циклів. Дослідницький цикл, як правило, складається з таких основних етапів:

- постановка завдання;
- цілепокладання і планування;
- усвідомлення завдання, його складових, що описують сферу дослідження;
- добір даних;
- визначення наочних галузей, що описують сферу дослідження;
- добір словників, тезаурусів, класифікаторів й інших матеріалів, що описують наочні галузі;
- формування баз даних (бібліотек);
- опрацювання даних – аналіз за допомогою відповідних методів й інструментів;
- структуризація зібраних даних (вибір формату й носіїв);
- опрацювання й підготовка матеріалу;
- визначення тематик доменних картриджів, таксономій, класифікацій і їх комбінацій для розробки ЛР;
- розробка і тестування ЛР;
- визначення кола користувачів ЛР і баз даних;
- використання інформації в процесі прийняття і виконання рішень.

Використання запропонованого методу побудови моделі знань навчальної дисципліни дозволяє також урізноманітнити процес навчання і зробити його більш персоналізованим. Це досягається за рахунок того, що учень має можливість використовувати свій власний досвід, будувати свої моделі знань, експериментувати з різними варіантами розв'язування задач, активно використовуючи функціональність комп'ютерного тезауруса. Методика навчання системам понять у середовищі тезауруса деякою мірою є ілюстрацією принципу навчання на основі дослідження й відкриття.

Висновки. Проведення досліджень у вказаному напрямку передбачає вирішення актуальних проблем підвищення ефективності використання освітянських інформаційних ресурсів на основі застосування сучасних інформаційно-комунікаційних технологій за допомогою засобів дистанційного доступу до розподілених систем знань. Інноваційною метою є створення онтологічних описів і моделювання явищ, які є об'єктами вивчення в навчальному процесі, що стає одночасно засобом засвоєння методології наукового пошуку, інваріантного до змісту предметних галузей комп'ютерного аналізу й імітації.

Список використаних джерел

1. *Palagin A.* Informational model of natural language processing. / Palagin A., Gladun V., Petrenko N., Velychko V., Sevruk A., Mikhailiuk A. // International Journal "Information Technologies and Knowledge". –2008. – Vol. 2. – Pp. 5–6.
2. *Величко В.* Автоматизированное создание тезауруса терминов предметной области для локальных поисковых систем / Величко В., Волошин П., Свитла С. // “Knowledge – Dialogue – Solution” International Book Series “INFORMATION SCIENCE & COMPUTING”, Number 15. – FOI ITHEA Sofia, Bulgaria. – 2009. – Pp. 24–31.
3. *Гладун В. П.* Процессы формирования новых знаний / В. П. Гладун. – София : СД "Педагог 6", 1994. – 192 с.
4. *Гладун В. П.* Конспектирование естественно-языковых текстов / В. П. Гладун, В. Ю. Величко // Proceedings of the XI-th International Conference “Knowledge-Dialogue-Solution”(KDS'2005). – Varna, Bulgaria, 2005. – Vol. 2.– Pp. 344–347.
5. *Добров Б. В., Лукашевич Н.В., Невзорова О.А., Федун Б.Е.* Методы и средства автоматизированного проектирования прикладной онтологии / Добров Б. В., Лукашевич Н. В., Невзорова О. А., Федун Б. Е. // Изв. РАН. Теория и системы управления. – М., 2004. – № 2. – С. 58–68.
6. Засоби доступу до джерел знань, та їх використання в учбовому процесі педагогічного навчального закладу : метод. рекомендації / за ред. члена-кореспондента НАН України С. О. Довгого і канд. техн. наук О. Є. Стрижака. – АПН, ІОД, 2009. – 66 с.

7. *Найханова Л. В.* Основные аспекты построения онтологий верхнего уровня и предметной области / Л. В. Найханова // “Интернет-порталы: содержание и технологии”, ФГУ ГНИИ ИТТ “Информатика”. – М. : Просвещение, 2005. – С. 452–479.
8. *Ножов И. М.* Морфологическая и синтаксическая обработка текста (модели и программы) : дисс. ... на соиск. уч. степени канд. техн. наук [Электронный ресурс]. – М., 2003. – 140 с. – Режим доступа : <http://aot.ru/docs/Nozhov/msot.pdf>.
9. *Палагин А. В.* К проектированию онтологоуправляемой информационной системы с обработкой естественно-языковых объектов / Палагин А. В., Петренко Н. Г. // Математические машины и системы. – 2008. – № 2. – С. 14–23.
10. *Палагин А. В.* К вопросу системно-онтологической интеграции знаний предметной области / Палагин А. В., Петренко Н. Г. // Математические машины и системы. – 2007. – № 3, 4. – С. 63–75.
11. *Палагин О. В.* Архитектурно-онтологичні принципи розбудови інтелектуальних інформаційних систем / Палагин О. В., Петренко М. Г. // Математичні машини і системи. – 2006. – № 4. – С. 15–20.
12. *Палагин О. В.* Модель категоріального рівня мовно-онтологічної картини світу / Палагин О. В., Петренко М. Г. // Математичні машини й системи. – 2006. – № 3. – С. 91–104.
13. *Попов Э. В.* Общение с ЭВМ на естественном языке / Э В. Попов. – М. : Наука, 1982. – 360 с.
14. *Стрижак О. Є.* Комп'ютерні тезауруси як технологічна платформа створення авторських методик викладання предметних дисциплін / О. Є. Стрижак // Актуальні проблеми психології: Психологічна теорія і технологія навчання / за ред. С. Д. Максименка, М. Л. Смольсон. – К. : Вид-во НПУ імені М. П. Драгоманова, 2009. – Т. 8. – Вип. 6. – С. 259–266.
15. *Тетелец Я. Г.* Введение в общий синтаксис / Я. Г. Тетелец. – М. : РГГУ, 2001. – 798 с.

**КОМПЬЮТЕРНЫЕ ОНТОЛОГИИ – ТЕХНОЛОГИЧЕСКАЯ ОСНОВА
ФОРМИРОВАНИЯ ОБРАЗОВАТЕЛЬНЫХ ИНФОРМАЦИОННЫХ РЕСУРСОВ**
Стрижак Александр Евгеньевич, советник директора Института одаренного ребенка НАПН Украины, кандидат технических наук, г. Киев,

e-mail:stryzhak953@gmail.com

Демьяненко Валентина Борисовна, аспирантка Института информационных технологий и средств обучения НАПН Украины, руководитель отделения «Компьютерные науки» Киевской малой академии наук ученической молодежи, г. Киев, e-mail:demyanenko@i.ua

Аннотация

Описываются подходы, модели и средства формирования систем знаний с целью дальнейшего развития образовательных информационных ресурсов. Рассмотрен онтологический аспект проектирования баз знаний учебного назначения. Определяются онтологические модели, средства лингвистического, структурного и функционального анализа и построения описаний объектов учебных дисциплин. Проведение исследований в указанном направлении предусматривает решение актуальных проблем повышения эффективности использования образовательных информационных ресурсов на основе применения современных информационно-коммуникационных технологий с помощью средств дистанционного доступа к распределенным системам знаний.

Ключевые слова: информационная модель, онтология, морфология, тезаурус, лингвистический и структурный анализ, база знаний.

COMPUTER ONTOLOGY AS TECHNOLOGICAL BASIS OF EDUCATIONAL INFORMATION RESOURCES FORMATION

Alexander E. Stryzhak, PhD, the adviser of the Director of the Institute of Gifted Child of the National Academy of Pedagogical Sciences of Ukraine, Kyiv, e-mail:stryzhak953@gmail.com

Valentina B. Demyanenko, the post-graduate student of the Institute of Information Technology and Learning Tools of the National Academy of Pedagogical Sciences of Ukraine, the Head of the Department “Computer sciences” of Kyiv’s “Small Academy of Sciences of Student's Youth of Ukraine”, Kyiv, e-mail:demyanenko@i.ua

Rezume

Approaches, models and means of knowledge systems formation with objective of the further progress of educational information resources are described. The ontological aspect of designing of knowledge bases of educational purpose is considered. Ontological models, means of the linguistic, structural and functional analysis and construction of

objects descriptions of academic disciplines are determined. Carrying out of researches in the specified direction provides the decision of actual problems of increase of efficiency of use of educational information resources based on application of modern information and communication technologies by means of remote access to the distributed knowledge systems.

Keywords: informative model, ontology, morphology, thesaurus, linguistic and structural analysis, base of knowledge.

Матеріал надійшов до редакції 11.03.2011 р.