

## ОСОБЛИВОСТІ ФОРМУВАННЯ ЕЛЕКТРОННОЇ ЕНЦИКЛОПЕДІЇ ВИЩОГО НАВЧАЛЬНОГО ЗАКЛАДУ

© Жежнич П.І., Гірняк М.Г., 2013

Наведено основні характеристики та кроки при формуванні електронної енциклопедії навчального закладу та її якісного функціонування. Проаналізовано доцільність екстракції знань з відкритих текстів для інформаційного наповнення та оцінювання якості енциклопедії.

**Ключові слова:** електронна енциклопедія, інформаційне наповнення, якість, екстракція знань, відкриті тексти, життєвий цикл.

**This paper dwells on the characteristic features and main steps in the formation of the university electronic encyclopedia and its qualitative performance. It is made up the analysis on the expediency of knowledge extraction from open texts for the content and encyclopedia quality evaluation.**

**Key words:** electronic encyclopedia, content, quality, knowledge extraction, open texts.

### Вступ

Упродовж останніх десятиліть спостерігається значний розвиток інформаційно-комунікативних технологій, що відображається у середовищі вищих навчальних закладів (ВНЗ). Свідченням цього є створення освітніх порталів, тренажерів, впровадження дистанційного навчання, проведення онлайн-семінарів, комп'ютерного тестування та розроблення електронних енциклопедій вищих навчальних закладів. Що ж стосується енциклопедії навчального закладу, то це – освітній ресурс, що інформує зацікавлених осіб (абітурієнтів, батьків, викладачів, студентів тощо) про освітні послуги, наукові надбання, дослідження, події, міжнародну співпрацю. Енциклопедія такого роду сприятиме підвищенню іміджу науково-дослідного закладу, слугуватиме своєрідною візитівкою [1].

### Постановка проблеми

Електронні енциклопедії надають довідкову інформацію щодо різних запитань. Отже, будь-який користувач, послуговуючись електронною енциклопедією вищого навчального закладу, отримує усю необхідну інформацію щодо діяльності установи. Тобто, слід розглянути основні особливості формування енциклопедії такого спрямування з огляду на відповідне функціонування та інформаційне наповнення. Про якість електронної енциклопедії свідчить доречність використання програмних засобів та вичерпний і достовірний контент.

### Мета роботи

Метою роботи є аналіз особливостей електронної енциклопедії вищого навчального закладу як інформаційної системи.

### Виклад основного матеріалу

Електронні енциклопедії відзначаються науково вивіреною інформацією із достовірних джерел із бібліографічними посиланнями. Для кращого візуального сприйняття та засвоєння інформації слід поєднувати подання матеріалу у вигляді діаграм, графіків, зображень, аудіо- та відеоінформації та високоякісного ілюстративного матеріалу. Для енциклопедії ВНЗ такий спектр контенту неабияк важливий, оскільки свідчитиме про його авторитет.

Розробляючи електронну енциклопедію вищого навчального закладу, слід розглянути життєвий цикл електронної енциклопедії як інформаційної системи, який складається із сукупності таких етапів:

- 1) аналіз вимог;
- 2) проектування;
- 3) реалізація / впровадження;
- 4) тестування і верифікація;
- 5) експлуатація.

Зупинимось детальніше на кожному з етапів, зосередившись на структурних особливостях електронної енциклопедії та відкритих текстах, з яких здійснюється екстракція знань для інформаційного наповнення енциклопедії.

Екстракція знань – це методологія, в основу якої покладено тріаду: “об’єкт, предмет і метод дослідження”: об’єктом є електронні тексти, предметом – комп’ютерні технології їх інтелектуальної обробки, основні методи – статистичний аналіз і теорія нечітких множин [2]. Для інформаційного наповнення електронної енциклопедії застосовуємо екстракцію знань із відкритих текстів, оскільки більшість інформаційних ресурсів, що містяться у веб-мережі (World Wide Web), – це неформалізовані тексти, які вимагають обробки за змістом. Екстракція знань уможливує порівняння, аналіз та синтез різномірної інформації, що міститься у відкритих текстах, доступних широкому загалу та призначених для зберігання, передавання та перетворення [2].

**1. Аналіз вимог.** Вимоги до інформаційного наповнення полягають у тому, що: а) джерелом інформаційного наповнення електронної енциклопедії вищого навчального закладу є відкриті тексти, які обробляються із застосуванням екстракції знань; б) інформаційне наповнення повинне містити логічний, послідовний, взаємозалежний виклад матеріалу із довідковим характером з наукововивірених джерел.

Під час екстракції знань з відкритих текстів можуть виникнути такі ускладнення [4; 5, с. 67]:

– *складність відкритого тексту:*

- складність мови відкритого тексту;
- якість тексту (залежить від грамотності автора);
- наявність / відсутність часткової структури документа;
- наявність апріорних знань про зміст тексту (н-д, резюме);
- наявність / відсутність стильової / жанрової однорідності текстів;

– *складність цільових знань (знань для екстракції):*

- різноманітність способів викладу необхідних знань у текстах;
- *відсутність цілісної системи знань через вибірккову екстракцію фрагментів;*
- *різна термінологія внаслідок екстракції знань з різних відкритих текстів;*
- *неадекватна модель (мова) для подання знань.*

**2. Проектування.** Етап проектування супроводжується:

- формулюванням тематичних розділів енциклопедії (класифікаторів);
- розробленням інформаційної моделі словника та статей електронної енциклопедії;
- побудовою архітектури системи електронної енциклопедії вищого навчального закладу, враховуючи потреби та зацікавлення користувачів усіх вікових категорій та сфер діяльності.

Електронна енциклопедія ВНЗ містить такі рубрикатори (класифікатори): структурованість університету; повсякденне життя закладу; людський ресурс; види діяльності (навчальні та наукові напрями / спеціальності); будівлі закладу; публікації про установу (газетні та інтернет-публікації) та додатково – словник спеціалізованих термінів для інформування користувачів із того чи іншого питання з можливістю переходу на статтю для детальнішого ознайомлення.

Отже, розроблення інформаційної моделі електронної енциклопедії вищого навчального закладу полягає у виділенні певних інформаційних сутностей із зазначених рубрикаторів (класифікаторів). Зокрема такими сутностями є: подія, діяльність, особа, підрозділ, будівля / споруда / ландшафт, документ і додатково – терміни. Кожна сутність характеризується набором

атрибутів (груп атрибутів), які у повному обсязі розкривають зміст статті із кожного класифікатора. На основі аналізу інформаційних сутностей та їхніх атрибутів структура різних за рубриками статей зводиться до єдиного шаблону, який становить узагальнену інформаційну модель статті електронної енциклопедії вищого навчального закладу.

З огляду на все вищезазначене, архітектура системи формування електронної енциклопедії складається у комплексі із таких підсистем (рис. 1):

- підсистема формування ключових слів за цільовою аудиторією (у вигляді відповідного словника);
- підсистема аналізу відкритих джерел інформації за термінологічним словником енциклопедії (класифікує інформаційні блоки відкритих текстів за ключовими словами);
- підсистема екстракції знань з відкритих ресурсів (виявляє знання з прокласифікованих інформаційних блоків);
- підсистема формування енциклопедичної статті (на основі вилучених знань);
- підсистема побудови перехресних посилань (встановлює зв'язок із зовнішніми та внутрішніми статтями / ресурсами).



Рис. 1. Система формування електронної енциклопедії

Загалом система формування електронної енциклопедії полягає у створенні нових (наступних) версій статей за допомогою зазначених підсистем. Тобто результатом функціонування цієї системи є нова стаття або нова (доповнена) версія наявної статті.

**3. Реалізація.** Електронна енциклопедія містить детальну систему класифікаторів (рубрикаторів) та систему внутрішнього пошуку – словник спеціалізованих термінів. Побудова електронної енциклопедії розпочинається із формування словника спеціалізованих термінів, при цьому кожен термін супроводжується створенням окремої енциклопедичної статті.

Проектування електронної енциклопедії – це автоматизований процес. Розробленням класифікаторів та пошукової системи займається людина-експерт, а комп'ютер вибирає необхідний матеріал для енциклопедичного інформаційного наповнення з виокремленням основних елементів цього матеріалу завдяки екстракції знань з відкритих текстів.

Розглядаючи процес екстракції знань з відкритих текстів, слід проаналізувати систему екстракції знань. Для інформаційного наповнення електронної енциклопедії ВНЗ першочерговим завданням є розроблення системи екстракції знань або ж використання існуючої.

Загалом система екстракції знань з відкритих текстів передбачає таку послідовність етапів:

- підготовка тексту – розглядають документ із вилученням тексту та інформації про структуру та стильове оформлення тексту; визначення мови та кодування тексту (при необхідності – перекодування); незначне коригування помилок або помилок оптичного розпізнавання;
- графематичний аналіз – перетворення тексту у послідовність токенів – слів та інших символічних послідовностей, що утворюють закінчений мовний об'єкт, та їх класифікація; пошук можливих меж речень;

- морфологічний аналіз – визначення частин мови та морфологічних атрибутів слів тексту;
- постморфологічний аналіз – в окремих випадках морфологічного аналізу, що супроводжується створенням спеціальних евристичних методів; сюди відносять зняття морфологічної омонімії;
- повний або частковий синтактико-семантичний аналіз – встановлення синтаксичних зв'язків у реченні; їх синтаксична та семантична класифікація;
- вирішення кореферентності – встановлення преференційної тотожності різних висловлювань про одну сутність;
- екстракція знань – пошук у тексті цільової інформації та її перетворення на структуровану форму; для флективних мов (зокрема української) додатково – нормалізація;
- уніфікація та ототожнення вилучених знань – зведення вилучених знань до єдиної концептуальної моделі; перехід від текстового представлення до змістового [3].

**4. Тестування і верифікація.** Етап тестування супроводжується перевіркою на якість інформаційного наповнення електронної енциклопедії. Критерії оцінювання цієї якості полягають у такому:

- подання статей у межах кожного класифікатора (рубрикатора);
- зв'язність статей;
- повнота та актуальність інформації.

**5. Експлуатація.** Розміщення електронної енциклопедії навчального закладу у веб-мережі, підтримка та розвиток за допомогою постійного редагування та оновлення задля повноти та актуальності інформаційного наповнення.

### Висновки

Отже, формування електронної енциклопедії ВНЗ відіграватиме неабияку роль для зацікавленого кола користувачів та сприятиме потраплянню до національних та глобальних рейтингів ВНЗ завдяки інтернет-присутності. Подальшого вивчення потребують питання розроблення методу підвищення якості електронної енциклопедії та автоматизації процесу наповнення електронної енциклопедії на основі екстракції знань з відкритих текстів.

1. Осадчий В.В. *Передумови та технології створення освітніх Інтернет-ресурсів // Проблеми інженерно-педагогічної освіти.* – Харків: Українська інженерно-педагогічна академія, 2009. – № 22–23. – С. 162–170. 2. Горькова В.И. *Информетрия (количественные методы в научно-технической информации) // Итоги науки и техники. Сер. Информатика. Т. 10.* – М.: ВИНТИ, 1988. – 328 с. 3. Жежнич П.І., Гірняк М.Г., *Формування електронної енциклопедії за допомогою екстракції знань з відкритих текстів // Інформація, комунікація, суспільство: Мат-ли – I Міжнар. наук. конф. ІКС-2012, 25-28 квітня 2012р., м. Львів.* – С. 32-33. 4. Куршев А.П., Сулейманова Е.А., Трофимов И.В., *Роль знань в системах извлечения информации из текстов // Программные системы: теория и приложения.* – М., 2012. – № 3 (12). – С. 57–70. 5. Гаврилова Т.А. *Базы знаний интеллектуальных систем / Т.А. Гаврилова, В.Ф. Хорошевский.* – СПб: Питер, 2000. – 384 с. 6. Жежнич П.І., Гірняк М.Г. *Архітектура системи формування електронної енциклопедії на основі відкритих текстів // Вісник Нац. ун-ту “Львівська політехніка”. Комп’ютерні системи проектування. Теорія і практика.* – 2012. – №747. – С. 79.