

УДК 004.5:81'374.82



## ИНСТРУМЕНТАЛЬНАЯ ЛЕКСИКОГРАФИЧЕСКАЯ СИСТЕМА В ОБЛАСТИ ФИЗИКИ МАГНИТНЫХ ЯВЛЕНИЙ

С.С. Дикарева<sup>1</sup>, В.Н. Бержанский<sup>2</sup>, С.Н. Полулях<sup>3</sup>, Е.В. Потапова<sup>4</sup><sup>1</sup>ТНУ им. В.И.Вернадского, Центр когнитивной и прикладной лингвистики, Симферополь, Украина<sup>2</sup>Таврический национальный университет им. В.И.Вернадского, Симферополь, Украина<sup>3</sup>Таврический национальный университет им. В.И.Вернадского, Симферополь, Украина<sup>4</sup>ТНУ им. В.И.Вернадского, Центр когнитивной и прикладной лингвистики, Симферополь, Украина

В статье представлены результаты работы по созданию редактора лексикографической системы трехязычного (русский-украинский-английский) словаря терминов по физике магнитных явлений. Разработана оригинальная лексикографическая параметризация термина. Лексикографическая база данных насчитывает более 900 словарных статей. Редактор лексикографической системы имеет функцию публикации содержимого лексикографической базы в виде HTML-сайта.

ЛЕКСИКОГРАФИЧЕСКАЯ СИСТЕМА, ЭЛЕКТРОННЫЙ СЛОВАРЬ, БАЗЫ ДАННЫХ, ПРОГРАММНЫЙ ИНТЕРФЕЙС

### Введение

Электронный русско-украинско-английский «Словник термінів з фізики магнітних явищ» создан группой авторов совместно с Институтом магнетизма НАН и МОН Украины (г.Киев) в рамках «Державної програми розвитку й функціонування української мови на 2004-2010 роки». В основу данного проекта положен мультидисциплинарный подход к представлению научных знаний по физике магнитных явлений в электронных лексикографических системах [1, 2]. Принципы создания электронных лексикографических систем описаны в цикле работ В.А.Широкова [3, 4].

Проект направлен на систематизацию украинской физической терминологии на основе методов когнитивной и прикладной лингвистики [5, 2]. Последние десятилетия большое внимание уделяется проблемам терминологии и терминографии украинского языка. Создание многоязычной (русско-украинско-английской) лексикографической базы является актуальной задачей для многих предметных областей [6-8].

Целевой аудиторией являются студенты и преподаватели технических специальностей, научные сотрудники, технические переводчики.

Лексикографическая база данных на данный момент насчитывает более 900 словарных статей (эквиваленты на русском, украинском и английском языке, дефиниции терминов). В основу положены два печатных источника [9, 10].

### 1. Постановка задачи

Целью данной работы было создание инструментальной среды для ввода и редактирования информации в трехязычной лексикографической базе. В концептуальной модели системы помимо редактирования языковых реестров заложены следующие специфические возможности: двуязыч-

ный (русский-украинский) интерфейс редактора; ввод дефиниций термина на трех языках; наличие вариантов дефиниции; независимое редактирование информации о библиографических источниках, персоналиях, разделах физики магнитных явлений (ФМЯ); формирование пользовательской версии словаря в автоматическом режиме.

### 2. Структура лексикографической базы данных

Лексикографические параметры термина в созданной лексикографической системе разделены на две группы:

1. Параметры, описывающие термин внутри определенного языка (орфографический стандарт, грамматические характеристики, синонимы, дефиниция, пример контекстного употребления).

2. Параметры, содержащие информацию, универсальную для всех трех языков (семантические связи термина — гипонимы, гиперонимы; ассоциативные связи; персоналии; раздел ФМЯ).

В структуре лексикографической базы данных (ЛБД) выделены три языковые группы таблиц: английская, русская (рис. 1) и украинская, связанные друг с другом через таблицу `tbl_term`. Информация, не соотносимая с конкретным языком (родовидовые связи термина, персоналии, раздел ФМЯ) образуют отдельную группу таблиц, связанную с терминами конкретного языка через таблицу `tbl_term` (рис. 2).

На рис. 1. `tbl_term` — это общая для трех языков таблица переходов, `tbl_term_rus` — русский языковой реестр, `tbl_term_rus_synonym` — таблица для записи пар индексов терминов-синонимов для русского языка, `tbl_term_rus_details` — контекстные примеры употребления термина, `tbl_term_rus_descn` — текстовое описание термина.

Как отмечалось, в лексикографическую параметризацию термина включены гипонимы и ги-

пероним для данного термина, которые отражают связи термина типа «род-вид». Кроме того, включены ассоциативно связанные термины (по отношению к данному), определяемые экспертами исходя из содержания предметной области - физики магнитных явлений. Перечисленные выше типы отношений, в отличие от отношения синонимии, не являются внутриязыковым явлением и в базе данных хранятся в общей для всех языков таблице отношений (рис. 2).

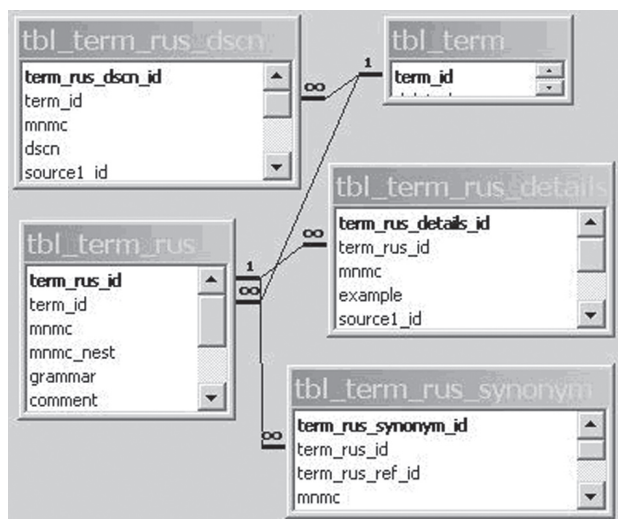


Рис. 1. Группа таблиц русского языка

На рис. 2 tbl\_term – это общая для трех языков таблица переходов, tbl\_term\_rus, tbl\_term\_ukr, tbl\_term\_eng - русский, украинский и английский языковые реестры. tbl\_relation\_term – таблица для хранения пар индексов терминов, находящихся в определенном отношении (relation\_type\_id): гипонимы, гипероним, ассоциативно связанный термин. tbl\_relation\_type – реестр типов отношений. Таблица tbl\_person\_term связывает термины с персоналиями, tbl\_subject\_term – с разделами ФМЯ.

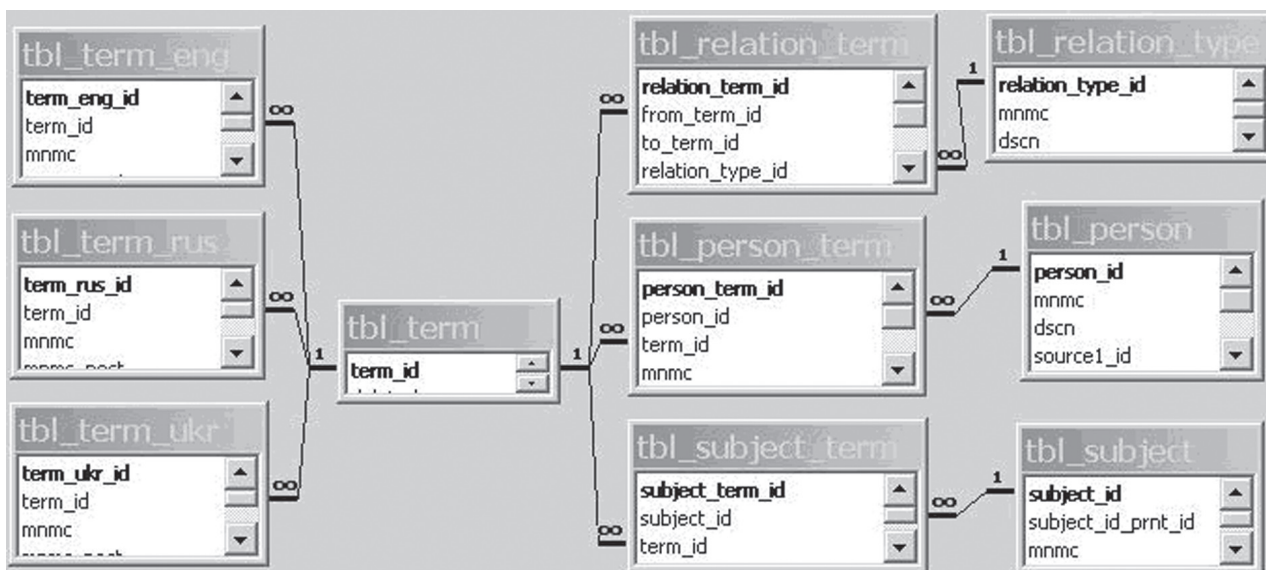


Рис. 2. Схема хранения связей термина в базе данных

### 3. Архитектура программного интерфейса редактора словаря

Работу с электронной лексикографической системой можно условно разделить на два режима: 1) работа редактора лексикографической базы – ввод данных в ЛБД и редактирование; 2) работа конечного пользователя словаря – поиск и просмотр информации, хранимой в лексикографической базе данных. Исходя из необходимости максимально упростить системные требования к пользовательскому компьютеру, оказалось целесообразно создать два интерфейса: редактор словаря (табл. 1) и версию словаря для конечного пользователя.

Редактор словаря реализован как локальное Windows-приложение (С#) с доступом к источнику данных – файлу словарной базы (ACCESS). Возможен удаленный доступ.

Таблица 1

Описание объектов и процедур редактора лексикографической базы данных

Группы объектов и процедуры	Назначение, функциональность
Управляющие элементы и объекты форм	Вывод информации ЛБД на экран, ввод запросов редактора, ввод и редактирование информации редактором-экспертом
Промежуточные объекты	Предварительная обработка массивов данных: хранения и преобразования вводимых и выводимых данных, формирования хеш-таблиц. Формирование SQL-запросов
Объекты и процедуры ввода/вывода	Установление связи с базой данных, редактирование и ввод новой информации по запросам (см. пред. шаг) в базу данных
Процедура генерации HTML кода	Генерация для каждого термина отдельной HTML страницы, заполненной информацией из ЛБД, с необходимыми гиперссылками для связи с другими страницами

Ступенчатая архитектура доступа к словарной базе данных (рис.3) имеет преимущество перед прямым доступом. Она позволяет модернизировать интерфейс редактора без вмешательства в другие группы объектов (табл. 1). Объекты ввода/вывода представляют собой отдельную библиотеку, которая в дальнейшем будет включена в сетевую версию редактора лексикографической базы.

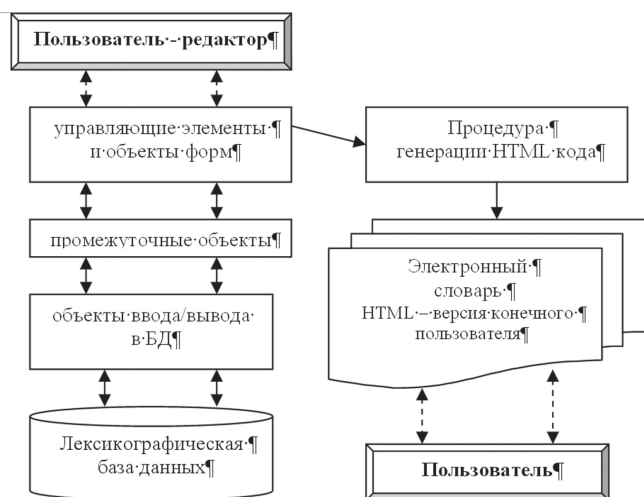


Рис. 3. Архитектура программного интерфейса редактора

Требования к компьютеру редактора словаря – операционная система Windows XP и последующих версий, Microsoft Office в конфигурации, включающей ACCESS.

Графический интерфейс (рис. 4.) редактора лексикографической базы разработан в соответствии с лексикографическими параметрами словаря. Меню приложения, панель инструментов, главный языковой реестр словаря (русский, украинский или английский – выбор зависит от языка запроса пользователя) присутствуют на экране в течение всего времени работы приложения. Также присутствует отображение искомого термина на трех языках (лексикографические параметры: орфографический стандарт, переводные эквиваленты).

Остальная информация группируется в отдельные вкладки следующим образом:

1. Описание (лексикографический параметр: дефиниция)
  - Текстовое описание.
  - Источник описания.
2. Примеры употребления
  - Пример употребления в тексте (отрывок из научной или научно-популярной статьи) (лексикографический параметр: контексты употребления).
  - Источник примера.
3. Дополнительно
  - Раздел ФМЯ.
  - Персоналии. Сведения о персоналиях, так или иначе связанных с данным термином.
4. Грамматика
  - Грамматическое описание термина в трех языках (лексикографический параметр: грамматикализация).

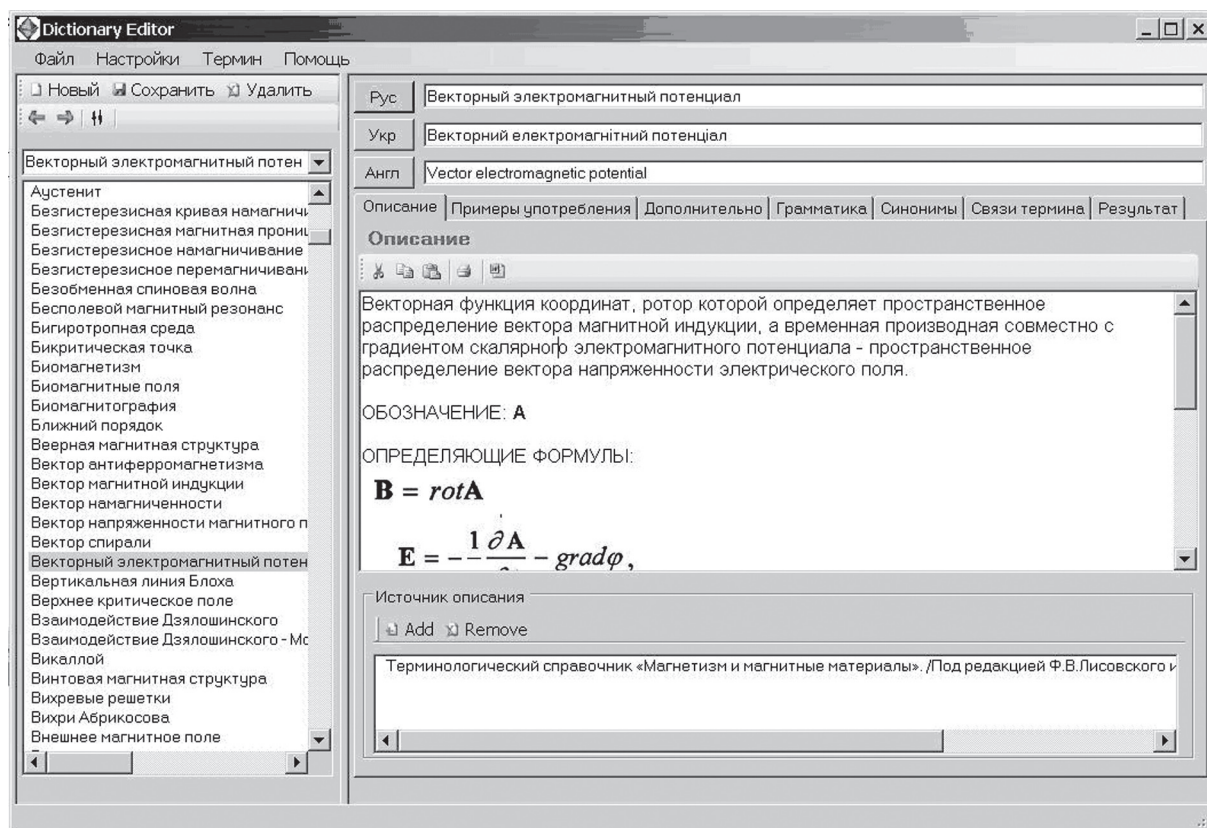


Рис. 4. Интерфейс эксперта. Вкладка «Описание»

- Комментарий (лексикографический параметр: ремарки).

5. Синонимы (лексикографический параметр: варианты и синонимы).

6. Связи термина – информация, необходимая для построения карты термина.

- Гипероним (лексикографический параметр: гипероним (родовое понятие)).

- Гипонимы (лексикографический параметр: гипонимы (видовые понятия)).

- Термины, ассоциативно связанные с данным термином (лексикографический параметр: ассоциативный ряд).

7. Результат – словарная статья термина (в формате HTML), построенная на основе информации, введенной в ЛБД на предыдущих вкладках.

В приложении предусмотрен выбор языка интерфейса пользователя: русский, украинский, английский. Он осуществляется с помощью команд Главного меню: **Настройки>Украинский интерфейс** (Русский, Английский там же). Полный перечень команд Главного меню и их назначение содержит табл. 2.

Таблица 2

Команды Главного меню

Команда меню	Назначение
Файл> База данных	В диалоговом окне указать путь к базе данных
Файл> Обновить	Обновить главный реестр терминов
Файл> Генерировать HTML	Запуск процедуры генерации набора HTML-страниц для создания версии словаря для конечного пользователя
Файл> Выход	Выход из программы
Настройки> Русский интерфейс	Выбор языка интерфейса
Настройки> Украинский интерфейс	Выбор языка интерфейса
Настройки> Английский интерфейс	Выбор языка интерфейса
Настройки> Источники	Открывает диалог для добавления или редактирования выходных данных источников описаний и текстовых примеров
Настройки> Персоналии	Открывает диалог для добавления или редактирования сведений о персоналиях
Настройки> Разделы ФМЯ	Открывает диалог для добавления или редактирования разделов ФМЯ
Термин> Новый...	Диалог для добавления нового термина в словарь
Термин> Сохранить	Сохраняет всю информацию для выбранного на данный момент термина
Термин> Удалить	Удаляет выбранный термин
Помощь> О программе	Демонстрирует сведения о проекте

На любом этапе работы над словарем содержимое словаря может быть опубликовано в электронном виде. Создание версии электронного терминологического словаря для конечного пользователя происходит по команде **Файл>Генерировать HTML** из меню редактора словаря. Запускается процедура, генерирующая HTML код (см. рис. 3) для отображения данных, находящихся в лексикографической базе на момент запуска процедуры. Вся информация лексикографической базы отображается в виде отдельных для каждого термина HTML-страниц и сохраняется в отдельную папку. К этому набору страниц были добавлены служебные файлы (собственно, **Интерфейс конечного пользователя**) для Главной страницы, Алфавитного указателя (на 3-х языках), Оглавление разделов ФМЯ и некоторые служебные файлы для оформления сайта и организации поиска.

#### 4. Интерфейс конечного пользователя

Интерфейс конечного пользователя представляет собой сайт, построенный на основе DHTML и Java-script, что позволяет минимизировать требования к компьютеру конечного пользователя. Пользовательская версия словаря может распространяться на лазерных дисках или размещаться в Интернете. Для установки пользовательской версии на локальном компьютере пользователя достаточно, чтобы на нем была установлена операционная система Windows.

В интерфейсе конечного пользователя предусмотрена группировка терминов по разделам ФМЯ и по Персоналиям (рис. 5). В отличие от редактора здесь сразу приводятся все три языковые реестры, причем в первую колонку помещается реестр на языке алфавитного указателя.

Словарная статья термина (рис. 6) помимо информации из базы данных содержит определенное количество гиперссылок, связывающих ее с другими словарными статьями, а также с источниками в Интернете (если был указан Интернет адрес). Каждая словарная статья открывается в отдельном окне, что дает возможность одновременно изучать информацию по нескольким терминам.

Часть словарной статьи, содержащая информацию о связях термина, представляет собой электронную интерактивную карту термина в табличном виде (рис. 7). Термины, связанные с данным, приводятся сразу на трех языках и представляют собой гиперссылки для быстрого перехода к словарной статье выбранного термина.

Карта термина отражает участок дерева концептов предметной области [2] и позволяет воспринимать не отдельные статичные образы, а смысловые схемы, логические связи, характер взаимодействия между предметами и явлениями. Графическое отображение карты термина находится в стадии разработки.

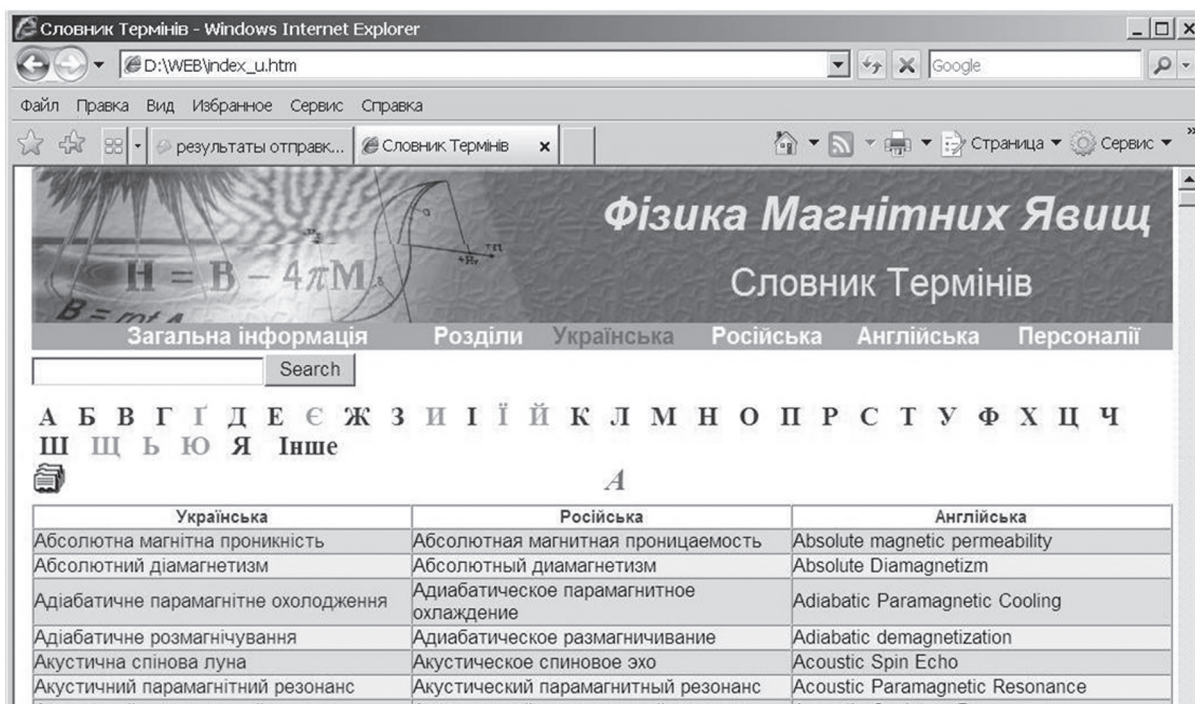


Рис. 5. Інтерфейс користувача – український алфавітний указатель

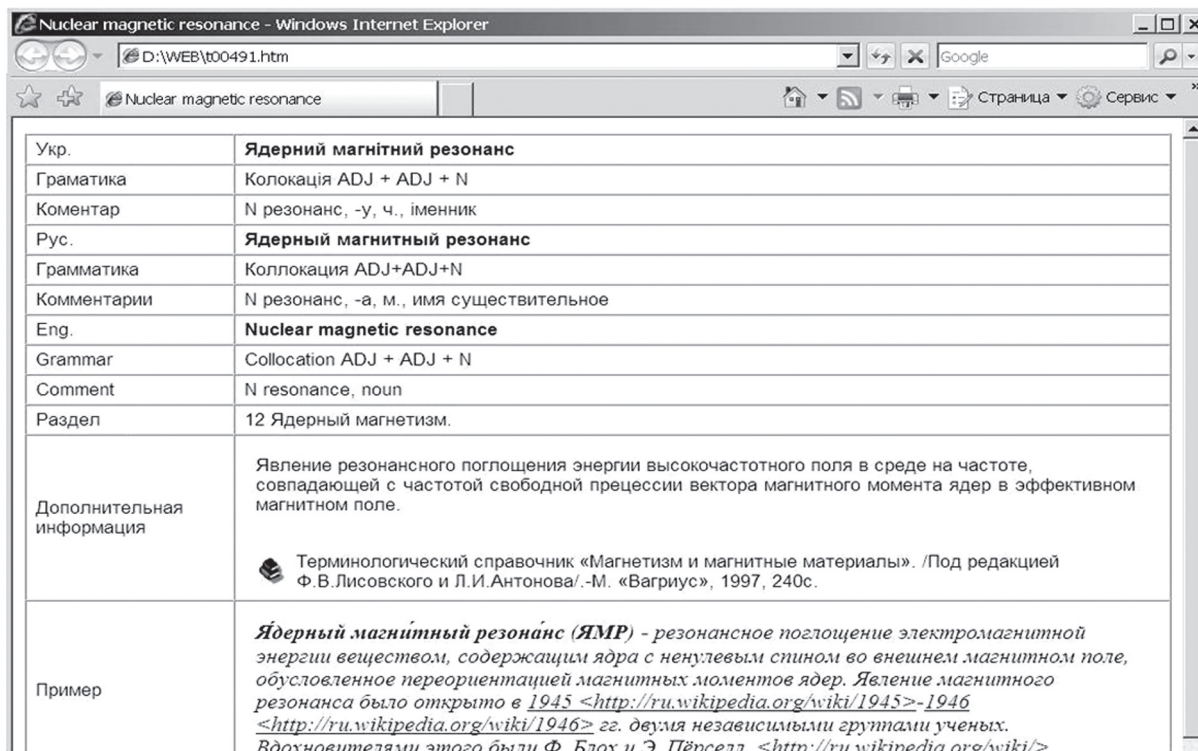


Рис. 6. Інтерфейс користувача. Словарна стаття терміна (частина 1)

### Заклучение

Разработанная инструментальная система (редактор) позволяет пополнять и редактировать содержимое лексикографической базы данных, а затем публиковать его в виде веб-сайта, построенного на основе HTML и Java-script. Этот сайт представляет собой пользовательскую версию словаря и может быть размещен в Интернете или записан на

лазерный диск. Редактор и пользовательская версия имеют двуязычный интерфейс (русский, украинский). Программный интерфейс и алгоритм описания терминов не зависят от предметной области и могут быть использованы для создания трехязычных словарей других предметных областей.

Авторы приносят глубокую благодарность директору Института магнетизма НАН и МОН Ук-

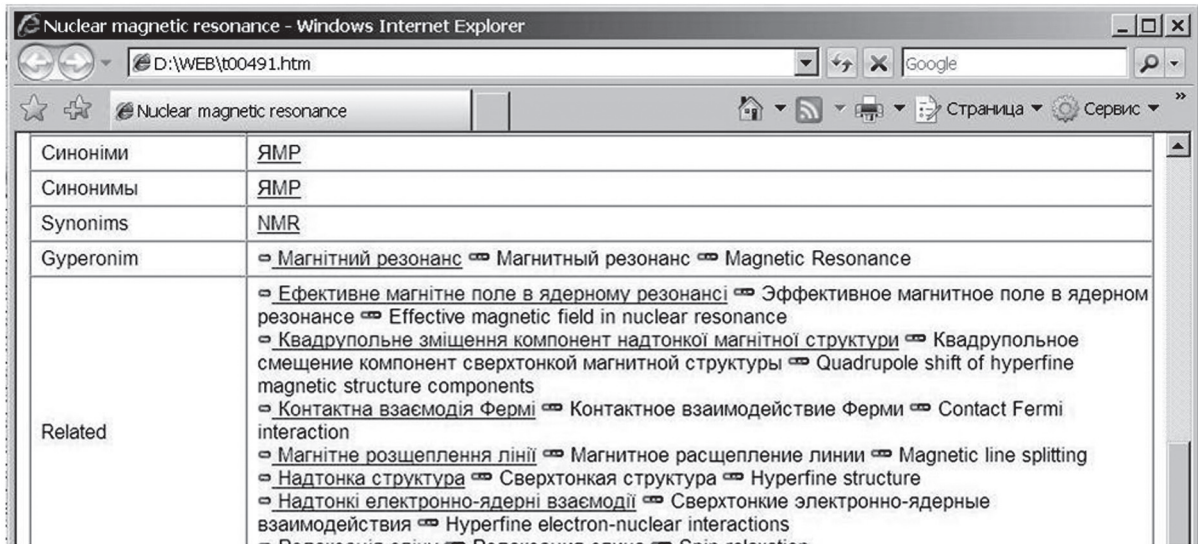


Рис. 7. Интерфейс пользователя. Словарная статья термина (часть 2 – карта термина)

раины академику НАН Украины В.Г.Барьяхтару и директору Украинского языково-информационного фонда НАН Украины член-корреспонденту В.А.Широкову за постоянный интерес к данному проекту и ценные советы.

**Список литературы:** 1. Бержанский, В.Н. Формирование онтологии на материале терминосистемы физики магнитных явлений [Текст] / В.Н.Бержанский, Е.В.Потапова, О.Ю. Салюк // Прикладна лінгвістика та лінгвістичні технології: MegaLing 2007: сб.науч. тр./ отв. ред. В.А.Широков – К.: Довіра, 2008. С.61-67. 2. Барьяхтар, В.Г. Ментальные пространства физики магнетизма в информационно-коммуникативном обществе знаний [Текст]/ В.Г. Барьяхтар, В.Н.Бержанский, С.С.Дикарева // MegaLing'2008 Горизонти прикладної лінгвістики та лінгвістичних технологій: тез. науч. конф. (вересень 2008) / отв. ред. В.А.Широков. – Сімферополь, ДИ-АЙ-ПИ, 2008. – С. 350 – 352. 3. Широков, В.А. Информационная теория лексикографических систем [Текст] / В.А.Широков. - Київ, Довіра, 1998. – 331с. 4. Широков, В.А. Элементы лексикографии [Текст] / В.А. Широков. – К.: Довіра, 2005. 5. Герд А.С. Научно-техническая лексикография [Текст] / А.С. Герд // Прикладное языкознание. – Спб, 1996. – 308 с. 6. Міжнародна наукова конференція “Проблеми української термінології” [Электронный ресурс] / Національний університет “Львівська політехніка”. – Режим доступа : [http://www.lp.edu.ua/tc.terminology/TK\\_konf.htm](http://www.lp.edu.ua/tc.terminology/TK_konf.htm). – 06.06.2009 г. – Загл. с экрана. 7. Технічний комітет стандартизації науково-технічної термінології [Электронный ресурс] / МОН України. – Режим доступа : <http://www.lp.edu.ua/tc.terminology> 25.03.2009 г. – Загл. с экрана. 8. Українська термінологія і сучасність [Текст]: збірник наукових праць. Випуск VII / Відп. ред. Л.О.Симоненко. – Київ: КНЕУ, 2007. – 207 с. 9. Физика твердого тела. Энциклопедический словарь: в 2 т. [Текст] / Под ред. В. Г. Барьяхтара. Т. 1.– Киев: Наукова думка, 1996. – 656 с., Т. 2. – Киев: Наукова думка, 1998. – 648 с. 10. Терминологический справочник «Магнетизм и магнитные

материалы» [Текст] / Под ред. Ф. В. Лисовского и Л. И. Антонова. – М.: «Вагриус», 1997. – 240 с.

*Поступила в редколлегию 06.10.2009*

УДК 004.5:81'374.82

**Інструментальна лексикографічна система в галузі фізики магнітних явищ** / С. С. Дікарева, В. Н. Бержанський, С. Н. Полулях, Є. В. Потапова // Біоніка інтелекту: наук.-техн. журнал. – 2009. – № 2 (71). – С. 48-53.

Розроблена інструментальна система (редактор) дозволяє поповнювати й редагувати вміст лексикографічної бази даних тримовного (російська-українська-англійська мови) словника термінів з фізики магнітних явищ, а потім публікувати вміст словника у вигляді веб-сайта, побудованого на основі HTML і Java-script. На основі методів когнітивної й прикладної лінгвістики розроблена оригінальна лексикографічна параметризація терміна. Програмний інтерфейс і алгоритм опису термінів не залежать від предметної галузі й можуть бути використані для створення тримовних словників інших предметних галузей.

Табл. 2. Іл. 7. Бібліогр.: 10 найм.

UDK 004.5:81'374.82

**Instrument lexicographical system in the field of physics of magnetic phenomena** /S.S. Dikareva, V.N. Berzhanskiy, S.N. Polulyakh, E.V. Potapova // Bionics of Intelligence: Sci. Mag. – 2009. – № 2 (71). – P. 48-53.

The developed instrument system (editor) makes it possible to supplement and to edit contents of the lexicographical data base of the three lingual (Russian-Ukrainian- English) dictionary of terms in physics of magnetic phenomena and then publish contents of the dictionary in a form of the HTML and Java-script Web site. On the basis of cognitive and applied linguistics methods the original lexico-graphical parametrization of term was developed. The program interface and the algorithm of the terms description does not depend on subject area and they can be used for creating the three lingual dictionaries of other subject areas.

Fig. 7. Tabl. 2. Ref.: 10 items.