

УДК 658.012.011.56



МЕТОД И ТЕХНОЛОГИЯ ПРЕДСТАВЛЕНИЯ ЭТИМОЛОГИЧЕСКОГО СЛОВАРЯ УКРАИНСКОГО ЯЗЫКА В ЦИФРОВОЙ СРЕДЕ

И. В. Остапова

Украинский языково-информационный фонд НАН Украины, г. Киев, Украина,
iros@zeos.net

На основе формальной модели лексикографической системы этимологического словаря украинского языка разработана технология построения инструментальной системы для поддержки функционирования словаря в цифровой среде. Основное внимание уделяется механизму языковой индексации словаря.

ЛЕКСИКОГРАФИЧЕСКАЯ СИСТЕМА, ЭТИМОЛОГИЧЕСКИЙ СЛОВАРЬ, ИНДЕКС СЛОВАРЯ, ЛЕКСИКОГРАФИЧЕСКАЯ БАЗА ДАННЫХ, ПОЛЬЗОВАТЕЛЬСКИЙ ИНТЕРФЕЙС

Введение

В последнее десятилетие словари в цифровой форме, предназначенные для компьютерного пользователя (как на CD-дисках, так и в режиме on-line), получили не только широкое распространение, но и имеют тенденцию занять лидирующее положение в сфере лексикографических изданий. Компьютерные словари предоставляют, по преимуществу, справочно-лингвистическую информацию в виде, адекватно отвечающем уровню существующего запроса. Сказанное, однако, не исключает возможности извлечения и концептуальной языковой информации, «скрытой», «зашифрованной» в лексикографических структурах. Это справедливо не только по отношению к обычным пользователям (для справочной информации), но и для лингвистов (для дальнейших исследований) [1–3].

Очевидное преимущество лексикографических трудов в цифровой форме — это эффективность поиска требуемой информации и возможность проведения вычислительных экспериментов. Особенно это актуально для многотомных словарных изданий. Основным же побудительным мотивом в деле внедрения новых информационных технологий в лексикографию служит необходимость вывести на новый уровень языковедческие исследования [1, 6]. Чрезвычайно долгий цикл создания и модификации фундаментальных бумажных словарей приводит к тому, что представленная в них информация устаревает к моменту выхода книжного издания. Кроме того, сложившаяся технология подготовки словаря к печати делает невозможной внедрение в словарные структуры новых достижений теоретической лексикографии. Немаловажное значение имеет и предоставляемая цифровой средой возможность многократного использования однажды сформированных лексикографических структур и массивов многими профессионалами: лингвистами, лингвотехнологами и издателями.

1. Постановка задачи

Этимологические словари — это специальные лингвистические справочные системы, которые содержат информацию о генетических связях (этимологии) слов определённого языка или группы родственных языков. Этимологические словари используются как справочные издания для широкого круга пользователей, так и в качестве инструмента для дальнейших этимологических исследований.

Этимологический словарь украинского языка (далее — ЭСУЯ) представляет собой фундаментальный лексикографический труд, который создаётся в рамках проекта формирования национальной словарной базы Украины [4]. Первый том был издан в 1982 году, выход шестого тома ожидается в конце 2009 года; седьмой том по замыслу авторов будет представлять индекс ко всему словарному массиву. Для этимологических словарей наиболее действенным поисковым инструментом является индекс по языковой принадлежности слов, с которыми устанавливается генетическая связь в каждой словарной статье.

На сегодняшний день (на материале 5-ти вышедших из печати томов) представлены уже 232 различных языка, что было установлено в результате создания нами компьютерной лексикографической базы данных ЭСУЯ (речь об этом пойдет ниже). Для каждого языка в словаре необходимо создать отдельный индекс с идентификацией всех точных текстовых локализаций каждого слова данного языка. Для всего массива словаря предполагаемая размерность индекса — около 120 тысяч единиц. Трудоемкость работы по построению такого индекса столь велика, что задача создания его в «ручном» режиме представляется технологически неоправданной. Поэтому и была поставлена задача разработки специальной цифровой лексикографической среды, адаптированной к структурам ЭСУЯ и ориентированной на создание моноязычного индекса в автоматическом режиме.

Цифровая среда представляет собой качественно новый уровень сервиса для исследовательской работы с лингвистической информацией, представленной в словарной форме. В первую очередь это относится к индексным системам. Под индексацией словаря мы понимаем набор формализованных правил и процедур, на основании которых можно получить информацию об определённых классах языковых фактов, зафиксированных в словаре. Реализуются эти правила в форме пользовательских интерфейсов. Следует, однако, учитывать тот факт, что эффективность построения индексных схем для цифрового словаря возможна только в среде, достаточно формализованной.

При выполнении работы по созданию цифровой версии ЭСУЯ использовались методы, которые уже были успешно опробованы для решения подобных задач, в частности, для создания компьютерной лексикографической базы данных нового толкового словаря украинского языка [6].

2. Структура лексикографической системы этимологического словаря украинского языка

Мы рассматриваем словарь как информационную систему особого типа — лексикографическую. Согласно теории лексикографических систем это абстрактный языково-информационный объект, ориентированный на реализацию комплексного информационного описания лексико-грамматических структур определённого языка или совокупности языков [6].

Архитектура системы отвечает стандартной трёхуровневой архитектуре информационных систем ANSI/SPARK, согласно которой в информационной системе выделяются концептуальный, внутренний и внешний уровни представления данных [5].

В качестве концептуальной модели используется лексикографическая модель данных [6]. Ниже мы приводим её в несколько упрощённом виде:

$$\{I_0(D), V(I_0(D)), \beta, \delta[\beta], Red[V(I_0(D))]\},$$

где D — объект моделирования — словарь украинского языка; $I_0(D) = \{x_i\}$ множество реестровых единиц словаря, в теории лексикографических систем его принято называть множеством *элементарных информационных единиц*; $V(I_0(D))$ — множество описаний (интерпретаций) элементарных информационных единиц, то есть текстов словарных статей: $V(I_0(D)) = \{V(x_i)\}$ — словарная статья с заголовковым словом (реестровой единицей) x_i ; β — множество структурных элементов, которые абстрагируются в результате анализа текста словаря; $\delta[\beta]$ — структура, порождаемая на β оператором δ , так что ограничение $\delta[\beta]$ на $V(x)$ порождает микроструктуру $\delta(x)$ словарной статьи; $Red[V(I_0(D))]$ — механизм *рекурсивной редукции лексикографической системы*. Он

даёт возможность последовательно выявлять всё более тонкие детали структуры лексикографической системы.

Изложенная модель представляет концептуальный уровень архитектуры и строится на основе анализа полиграфической версии ЭСУЯ: полиграфического оформления, организации текстов словарных статей, их структурных элементов, интерпретируемых как идентификаторы соответствующих элементов лексикографических структур β и $\delta(x)$.

В качестве базового структурного элемента лексикографической системы ЭСУЯ мы определяем *этимологический класс*, который представляет собой блок линейного текста словарной статьи, вычленяемый по формальным признакам: структурная единица идентифицируется как этимологический класс, если в тексте словарной статьи можно выявить уникальные знаковые последовательности, используемые в качестве разделителей. Для ЭСУЯ нами выделены следующие типы этимологических классов: *класс реестрового слова* (обозначается *HEAD*), *класс дериватов* (*DER*), *класс славянских соответствий* (*SLAV*), *языковой класс* (*LANG*), *библиографический класс* (*BIBL*), *класс ссылок* (*LINK*). Каждый из этих классов имеет собственную текстовую структуру, что дает возможность построения процедуры идентификации типа каждого этимологического класса в словарной статье по формальным признакам.

Класс реестрового слова и *языковой класс* составляют минимальную структуру словарной статьи. Этимологические классы других типов являются факультативными.

Проиллюстрируем изложенное на примере двух небольших словарных статей, достаточно репрезентативных с точки зрения структуры. Тексты приводим в форме, максимально приближенной к печатной версии.

Пример 1 (словарная статья с заголовковым словом **абетка**):

абетка, [абетло] Пі, абетний (заст.) «элементарный»;— власне українська назва азбуки, утворена за вимовою перших двох букв алфавіту (*а, бе*), очевидно, під впливом назв *азбука, алфавет* і п. *abecadło* «тс.» (від вимови перших трьох букв *а, бе, се*).— Sadn. — Aitz. VWb. I 42.— Пор. **азбука, алфавіт**.

Этимологические классы для словарной статьи **абетка** (тексты классов подаются в угловых скобках):

HEAD ≡ <абетка>

DER ≡ <[абетло] Пі, абетний (заст.) «элементарный»>

LANG ≡ <власне українська назва азбуки, утворена за вимовою перших двох букв алфавіту (*а, бе*),

очевидно, під впливом назв *азбука*, *альфабет* і п. abecadlo «тс.» (від вимови перших трьох букв а, be, ce)>

BIBL ≡ <Sadn. — Aitz. VWb. I 42>

LINK ≡ <Пор. **азбука**, **алфавіт**>

Пример 2 (словарная стаття с заголовковым словом **ва-банк**):

ва-банк (картярський термін) «ставка, що дорівнює всьому банку», (перен.) «ризкуючи всім»; — р. болг. *ва-банк*, п. *va banque*, ч. *vabank*, *vabanque*, слц. *vabank*, схв. *ва банк*; — запозичення з французької мови; фр. *va banque* (букв.) «іде банк» складається з дієслова *va* (3 ос. одн.), яке походить від лат. *vadit* «іде», спорідненого з двн. *watan* «переходити вбрід», дісл. *Vada*, дангл. *wadan* «тс.», і іменника *banque* «банк». — СІС 116; Шанский ЭСРЯ I 3,3; ССРЛЯ 2, 19; Dauzat 26, 737; Walde—Hofm. II 723—724. — Див. ще **банк**, **вада**.

Этимологические классы для словарной статьи **ва-банк**:

HEAD ≡ <**ва-банк** (картярський термін) «ставка, що дорівнює всьому банку», (перен.) «ризкуючи всім»>

SLAVIA ≡ <р. болг. *ва-банк*, п. *va banque*, ч. *vabank*, *vabanque*, слц. *vabank*, схв. *ва банк*>

LANG ≡ <запозичення з французької мови; фр. *va banque* (букв.) «іде банк» складається з дієслова *va* (3 ос. одн.), яке походить від лат. *vadit* «іде», спорідненого з двн. *watan* «переходити вбрід», дісл. *Vada*, дангл. *wadan* «тс.», і іменника *banque* «банк»>

BIBL ≡ <СІС 116; Шанский ЭСРЯ I 3,3; ССРЛЯ 2, 19; Dauzat 26, 737; Walde—Hofm. II 723—724>

*LINK*₁ ≡ <Див. ще **банк**, **вада**>

Приведём пример словарной статьи, минимальной как структурно, так и содержательно:

[**андріак**] «опій»; — походження неясне.

Пример 3 (этимологические классы для словарной статьи **андріак**):

HEAD ≡ <[**андріак**] «опій»>

LANG ≡ <походження неясне>

В тексте каждого этимологического класса устанавливаются связи реестрового слова с определёнными словами других языков. Все эти слова, включая реестровые, мы называем *этимонами*. При анализе текстов этимологических классов было выявлено восемь параметров, посредством которых описываются этимоны: *маркер языковой принадлежности* (обозначим P_L), *ремарка к маркеру языковой принадлежности* (P_{RL}), *знаковое представление этимона* (P_A), *принадлежность к диалектной лексике* (P_D), *маркер омонимии* (P_O), *толкование* (P_S), *ремарка* (P_R), *библиография* (P_B). Мы перечислили параметры в том порядке, в котором они, как правило, следуют в тексте соответствующего этимологического класса. Два параметра являются обязательными: это P_L (*маркер языко-*

вой принадлежности) и P_A (*знаковое представление этимона*). Эти два параметра обеспечивают уникальность каждого этимона словарной статьи: этимоны с одинаковой знаковой формой могут иметь разную языковую принадлежность, или этимоны с одинаковой языковой принадлежностью могут иметь разные знаковые формы. Остальные параметры — факультативные. Для каждого параметра определена формальная процедура, которая позволяет вычленить соответствующий параметр из текста для каждого этимологического класса.

Набор параметров $\{P_L, P_{RL}, P_A, P_D, P_O, P_S, P_R, P_B\}$ мы называем *этимон-структурой* и обозначаем $ETYM(e_i)$, где e_i — соответствующий этимон; индекс i — порядковый номер данного этимона в тексте. Порядок следования параметров в этимон-структуре полагаем не существенным.

Не все параметры актуальны для каждого этимологического класса. Текст, идентифицируемый как этимологический класс, использует свое подмножество параметров; не каждый этимон обязан описываться полным набором параметров. Однако для достижения структурной однородности для каждого класса строится один тип этимон-структуры; если определённый параметр не задействован или не может быть выделен по формальным признакам, то его значению соответствует пустая строка текста. Этимон-структура строится только в том случае, если удалось вычленить P_A . Формально мы полагаем, что каждому этимологическому классу соответствует этимон-структура. Если языковой класс не имеет ни одного этимона (или не удалось его выявить формальной процедурой), то мы считаем его вырожденным этимологическим классом и ему соответствует пустая этимон-структура. Примером такого класса служит языковой класс для словарной статьи из примера 3.

Проиллюстрируем этимон-структуры на примерах текстов этимологических классов:

Пример 4 (этимон-структуры для класса реестрового слова):

HEAD (**абетка**) ≡ <абетка>

$ETYM(e_1) \equiv \{P_L = \text{<укр.>}, P_A = \text{<абетка>}\}$

HEAD (**ва-банк**) ≡ <**ва-банк** (картярський термін) «ставка, що дорівнює всьому банку», (перен.) «ризкуючи всім»>

$ETYM(e_1) \equiv \{P_L = \text{<укр.>}, P_A = \text{<ва-банк>}, P_R = \text{<(картярський термін)>}, P_S = \text{<«ставка, що дорівнює всьому банку»>}\}$

Пример 5 (этимон-структуры для класса дериватов):

DER (**абетка**) ≡ <[*абетло*] Пі, *абетний* (заст.) «элементарный»>

$ETYM(e_1) \equiv \{P_L = \text{<укр.>}, P_A = \text{<абетло>}, P_D = 1, P_B = \text{<Пі>}\}$

$ETYM(e_2) \equiv \{<P_L = \text{<укр.>}, P_A = \text{<абетний>}, P_R = \text{<(заст.)>}, P_S = \text{<«элементарный»>}\}$

Параметр языковой принадлежности принимает значение «укр.» (украинский) по умолчанию, так как в этих текстах устанавливаются связи только со словами украинского языка. Параметр омонимии P_O для этимона e_1 принимает значение 1 (пример 5), так как квадратные скобки указывают на принадлежность слова к диалектной лексике. По умолчанию для всех этимонов значение этого параметра 0.

Пример 6 (этимон-структуры для класса славянских соответствий):

SLAV(ва-банк) $\equiv$ <р. болг. *ва-банк*, п. *va banque*, ч. *vabank*, *vabanque*, слц. *vabank*, схв. *ва банк*>

В тексте этого этимологического класса устанавливается связь со словами русского (р.), болгарского (болг.), польского (п.), чешского (ч.), словацкого (слц.) и сербскохорватского (схв.) языков.

$ETYM(e_1) \equiv \{<P_L = <р.>, P_A = <ва-банк>\}$

$ETYM(e_2) \equiv \{<P_L = <болг.>, P_A = <ва-банк>\}$

$ETYM(e_3) \equiv \{<P_L = <п.>, P_A = <va banque>\}$

$ETYM(e_4) \equiv \{<P_L = <ч.>, P_A = <vabank>\}$

$ETYM(e_5) \equiv \{<P_L = <ч.>, P_A = <vabanque>\}$

$ETYM(e_6) \equiv \{<P_L = <слц.>, P_A = <vabank>\}$

$ETYM(e_7) \equiv \{<P_L = <схв.>, P_A = <>\}$

Пример 7 (этимон-структуры для языкового класса):

LANG(ва-банк) $\equiv$ <запозичення з французької мови; фр. *va banque* (букв.) «іде банк» складається з дієслова *va* (3 ос. одн.), яке походить від лат. *vadit* «іде», спорідненого з двн. *watan* «переходити вбхід», дісл. *Vada*, дангл. *wadan* «тс.», і іменника *banque* «банк»>

В приведенном тексте устанавливаются связи со словами французского (фр.), латинского (лат.), древне немецкого (двн.), древнеисландского (дісл.) и древнеанглийского (дангл.) языков.

$ETYM(e_1) \equiv \{<P_L = <фр.>, P_A = <va banque>, P_S = <«іде банк»>, P_R = <(букв.)>\}$

$ETYM(e_2) \equiv \{<P_L = <фр.>, P_A = <va>, P_R = <(3 ос. одн.)>\}$

$ETYM(e_3) \equiv \{<P_L = <лат.>, P_A = <vadit>, P_S = <«іде»>\}$

$ETYM(e_4) \equiv \{<P_L = <двн.>, P_A = <watan>, P_S = <«переходити вбхід»>\}$

$ETYM(e_5) \equiv \{<P_L = <дісл.>, P_A = <Vada>\}$

$ETYM(e_6) \equiv \{<P_L = <дангл.>, P_A = <wadan>, P_S = <«тс.»>\}$

$ETYM(e_7) \equiv \{<P_L = <фр.>, P_A = <banque>, P_S = <«банк»>\}$

2. Парсинг этимологического словаря и создание его лексикографической базы данных

Основная проблема создания компьютерных словарей, исходя из их печатных версий, сводится к тому, каким образом сформировать соответствующую базу данных в автоматическом режиме не-

посредственно из текста словаря (парсинг). Опыт убеждает, что формирование нетривиальных лексикографических баз данных «вручную» из больших и сложных словарных текстов практически невозможно. Основной задачей парсинга является автоматическое выделение структурных элементов лексикографической системы непосредственно из текста словаря, поскольку именно они выполняют роль элементов лексикографической базы данных.

Парсинг ЭСУЯ был осуществлен следующим образом. Перед конверсией тексты всех томов были переведены в формат HTML и унифицированы как относительно структуры файлов, так и относительно знаковой системы. Разные тома словаря были подготовлены к печати различными издательскими технологиями. Первые три тома — в технологии монотайп, докомпьютерной. Поэтому печатные тексты сначала были отсканированы, распознаны программой FINEREADER, а затем вычитаны. Тексты 4-го и 5-го томов уже готовились в компьютерной издательской системе; для этого был разработан специальный набор компьютерных шрифтов, аналогичный используемому в 1, 2, 3 томах. Знаковая система словаря была унифицирована согласно кодировке UNICODE 3.0. Это позволило выполнить инвентаризацию символов алфавита для представления этимонов каждого языка.

Для связи между печатной и цифровой версиями словаря каждая словарная статья промаркирована следующим образом: номер тома, номер страницы начала текста, номер страницы конца текста.

Пример 8 (формат цифрового текста для статьи *ва-банк*; символ # используется для идентификации ударения):

@1 316 316

ва-ба#нк (картярський термін) «ставка, що дорівнює всьому банку», (перен.) «ризкуючи всім»; -- р. болг. <I>ва-ба#нк</I>, п. *va banque*, ч. *vabank*, *vabanque*, слц. *vabank*, схв. <I>ва банк</I>; -- запозичення з французької мови; фр. *va banque* (букв.) «іде банк» складається з дієслова *va* (3 ос. одн.), яке походить від лат. *vadit* «іде», спорідненого з двн. *watan* «переходити вбхід», дісл. *Vada*, дангл. *wadan* «тс.», і іменника *banque* «банк». -- C1C 116; Шанский ЭСРЯ I 3,3; ССРЯ 2, 19; Dauzat 26, 737; Walde--Hofm. II 723--724. -- Див. ще банк, ва#да.

В результате этих операций были получены специальным образом препарированные тексты томов этимологического словаря, полностью готовые для автоматической конверсии в лексикографическую базу данных.

Структура этой базы данных представляет ряд связанных таблиц, как это представлено на рис. 1.

Дадим краткую характеристику таблицам базы данных.

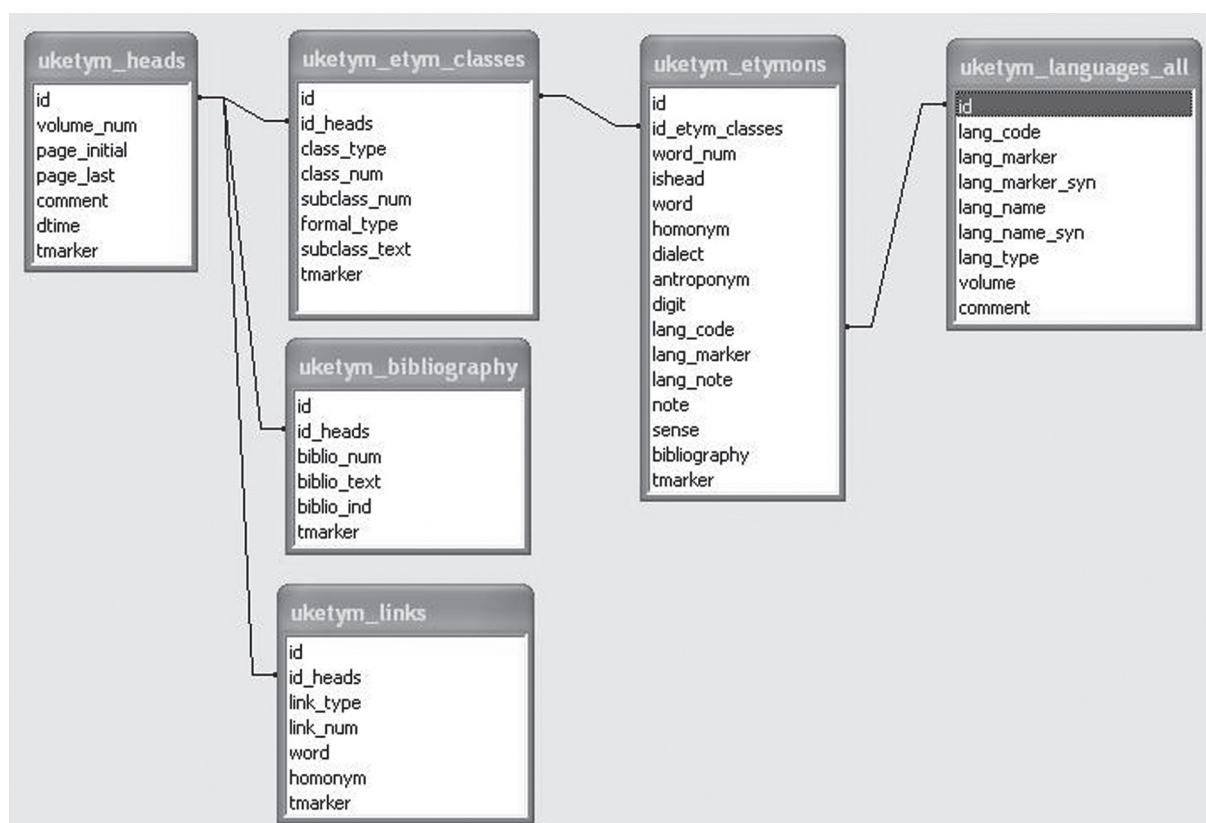


Рис. 1. Структура лексикографической базы данных этимологического словаря украинского языка

Таблица **uketym_heads** организует текст словарной статьи. Она хранит только идентификационный номер словарной статьи, номера тома, начальной и конечной страницы расположения текста словарной статьи в печатной версии.

Таблица языковых классов **uketym_etym_classes** хранит тексты словарных статей (за исключением библиографии и ссылок, которые организованы в отдельные таблицы).

Все межстатейные ссылки словаря организованы в таблицу **uketym_links**.

Библиография организована в таблице **uketym_bibliography**.

Таблица **uketym_etymons** содержит параметры этимон-структур.

В таблице **uketym_language_all** организована вся информация о языках, которые задействованы в словаре. На базе этой таблицы строятся пользовательские языковые регистры для индексирования. Приведём фрагмент этой таблицы:

id	Lang_code	lang_marker	lang_name	volume
127	127	п.	Польська	1
183	183	укр.	Українська	1
231	0	невизн.	невизначена	1

Как видим, в таблицу введен ещё один, «технологический», язык — неопределённый (мова «невизначена»). Это сделано из следующих соображений. Поскольку индексация производится автоматически, то не всегда согласно разработанно-

му алгоритму можно однозначно определить язык. В этом случае этимон маркируется неопределённым языком. По данному признаку из базы данных можно выбрать все подобные записи (например, сформировав языковой регистр, состоящий только из технологического языка) и затем произвести идентификацию языков, используя средства редактирования, предусмотренные в системе.

4. Функциональность инструментального этимологического комплекса

Для поддержки цифровой версии словаря построен инструментальный комплекс, который обеспечивает следующие основные функции [7]:

- 1) автоматическую конверсию текстов этимологического словаря в компьютерную базу данных;
- 2) традиционный вход в систему по реестровому слову и отображение текста словарной статьи;
- 3) редактирование любого структурного элемента словарной статьи;
- 4) построение этимон-структуры для словарной статьи в ручном режиме;
- 4) автоматическое построение этимон-структуры для словарной статьи;
- 5) создание словарной статьи с определенной структурой.

На рис. 2 показано одно из окон редактирования словарной статьи. На левой панели словарная статья представлена в виде дерева структурных элементов. Для каждого этимологического класса

выводится упорядоченный список этимонов, тем самым графическими средствами визуализируется глубина этимологического исследования. С помощью кнопок на средней панели структурные элементы можно добавлять, удалять и менять порядок их следования. Функции кнопок варьируются в зависимости от выбранного структурного элемента. Так, например, кнопка «Додати» (добавить) при выборе этимона позволяет добавить только этимон. Для каждого структурного элемента разработано своё окно редактирования, которое отражает специфику этого элемента. Для каждого этимона выводится и текст соответствующего этимологического класса, однако с запретом его редактирования. Это даёт возможность верифицировать параметризацию этимона, выполненную автоматически.

Для автоматического построения языковых индексов разработан специальный инструментарий, который позволяет:

- 1) в интерактивном режиме формировать любое количество языковых регистров на множестве всех языков словаря;
- 2) задавать спектры индексации, учитывая структуру словарной статьи.

На рис. 3 показано диалоговое окно пользователя для формирования языкового регистра.

Левая панель предназначена для выбора уже сформированных регистров в качестве неизменя-

емых шаблонов. Правая панель используется для редактирования существующих и формирования новых регистров.

В нашем примере мы сформировали новый регистр следующим образом: задали его имя «регистр_п_укр»; после проверки на уникальность новый регистр был занесён в список языковых регистров; в окно на левой панели загрузили регистр «Усі мови словника» (все языки словаря), который включает все задействованные в словаре языки; из списка были выбраны последовательно украинский и польский языки и перенесены в список на правой панели. Регистр «Усі мови словаря» может использоваться только в качестве шаблона. Предполагается в дальнейшем создать набор шаблонов со значимыми именами, например «Славянские языки», «Романские языки» и тому подобное.

На рис. 4 показано окно главного пользовательского интерфейса словаря с индексом, построенным по сформированному регистру.

На левой панели из предложенного набора языков был выбран польский (возможен выбор как всех языков в регистре, так и определённого подмножества языков). На правой панели в реестровое окно выведен список всех этимонов, которые идентифицированы как слова польского языка.

В окно реестра могут быть также выведены заголовковые слова тех словарных статей, в которых зафиксированы этимологические связи с поль-

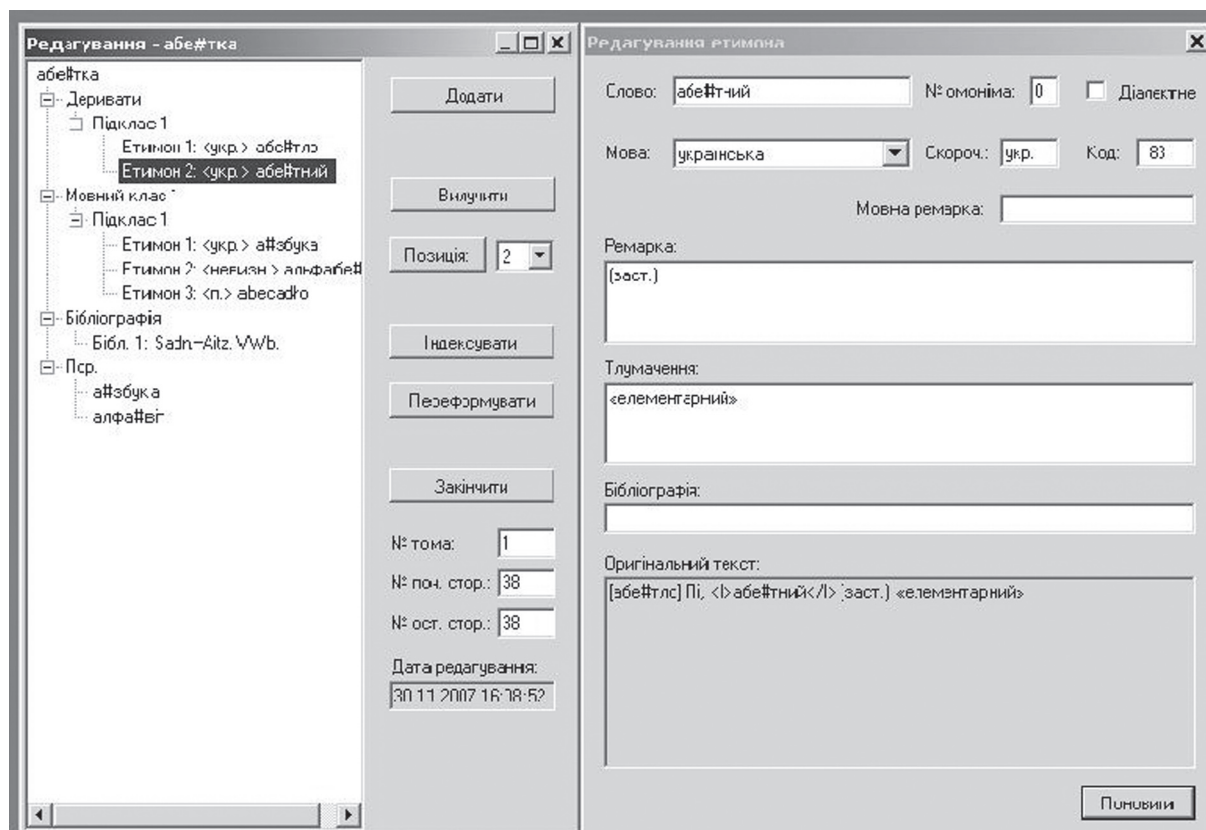


Рис. 2. Окно редактирования для словарной статьи с заголовковым словом **абетка**

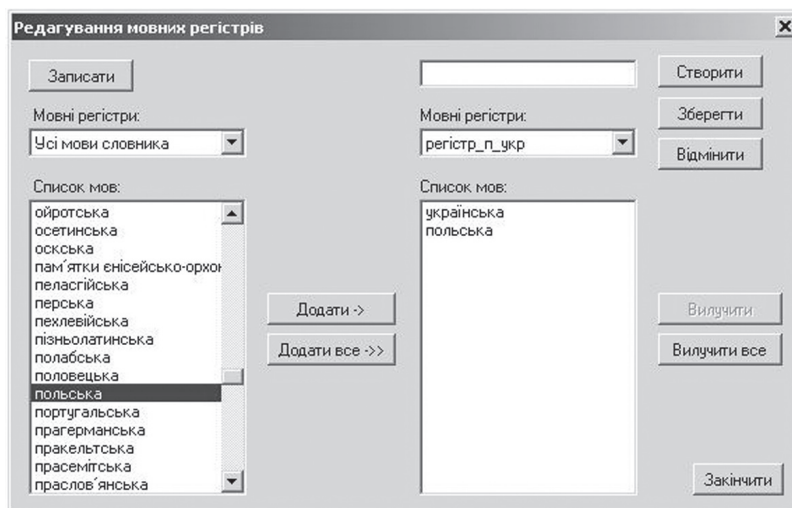


Рис. 3. Окно формирования языкового регистра

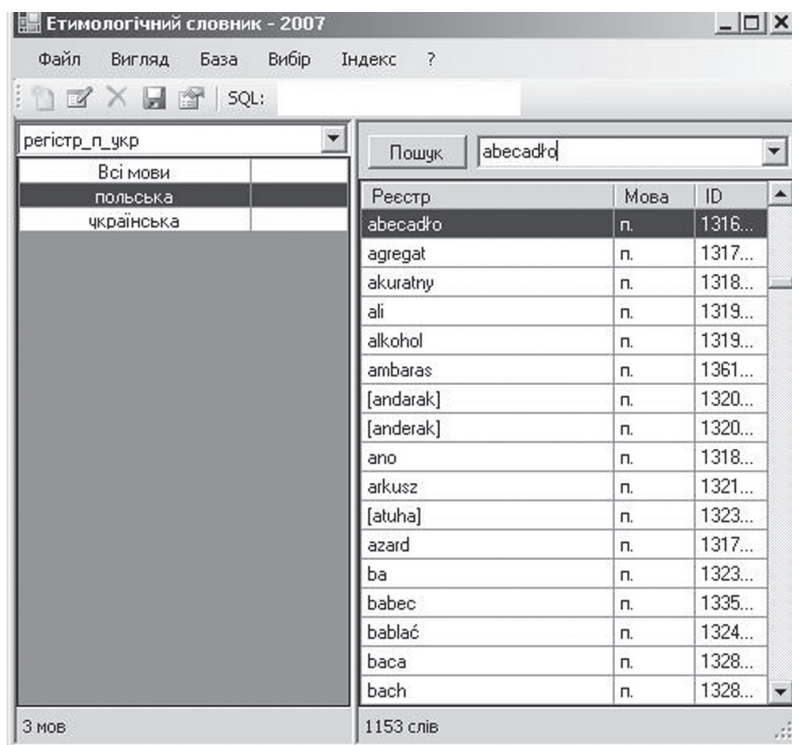


Рис. 4. Языковой индекс по заданному регистру

ским языком. Сформированный индекс по команде пользователя выводится в текстовый файл с указанием локализации каждого этимона.

Инструментальная система позволяет задать локализацию индексируемых элементов с точностью до структурного элемента — типа этимологического класса — словарной статьи (с помощью меню «Вибір» (Выбор) на верхней панели). В нашем случае был задан только *языковой класс*.

При активизации любого элемента реестра визуализируется текст словарной статьи.

Текст словарной статьи для вывода формируется из соответствующих полей базы данных. Полиграфическое оформление статьи практически сохранено полностью.

Выводы

Описанный метод представления этимологического словаря в цифровой среде дал возможность построить для совокупности словарных статей словаря соответствующую совокупность этимон-структур как формальных репрезентантов описаний генетических связей реестровых единиц. Все схемы индексации строятся только на основе этимон-структур. Такой подход обеспечивает как возможность построения структур, имплицитированных в текст словарных статей, так и отображения на цифровую среду аутентичного текста словаря, что делает цифровой словарь открытым для дальнейших интерпретаций.

Разработанные технологии и интерфейсы предлагаются как базовые для цифровых репрезентаций этимологических работ.

Список литературы: 1. Корпусна лінгвістика [Текст] / В. А. Широков, О. В. Бугаков, Т. О. Грязнухіна і др. ; відп. ред. В. А. Широков. — К. : Довіра, 2005. — 471 с. 2. *Григорович, Л. А.* Учебный словарь «NTC'American English Learner's Dictionary»: полиграфическая и электронная версии [Текст] / Л. А. Гончаров // Слово в словаре и дискурсе : Сб. науч. статей к 50-летию Харри Вальтера. — М. : ООО «Издательство “Элпис”», 2006. — 836 с. — С. 602–608. 3. *Большаков, И. А.* КроссЛексика — большой электронный словарь сочетаний и смысловых связей русских слов [Текст] / И. А. Большаков // Компьютерная лингвистика и интеллектуальные технологии : По материалам ежегодной Междунар. конф. «Диалог 2009» (Бекасово, 27–31 мая 2009 г.). Вып. 8(15). — М. : РГГУ, 2009. — XII, 620 с. — С. 45–50. 4. Етимологічний словник української мови : В 7 т. / Гол. ред. О. С. Мельничук. — Київ: Наукова думка, 1982–2006. Т. 1–5. 5. ANSI/X3/SPARK DBMS study group interim report. FDT-Bull. ACM SIGMOD. 1975. V. 7. № 2. 6. *Широков, В. А.* Елементи лексикографії [Текст] / В. А. Широков. — Київ: Довіра, 2005. — 304 с. 7. *Остапова, И. В.* Инструментальная лексикографическая система Этимологического словаря украинского языка [Текст] / И. В. Остапова, К. Н. Якименко // Прикладна лінгвістика та лінгвістичні технології : MegaLing-2007 : зб. наук. пр. — Київ: Довіра, 2008. — 511 с. — С. 276–291.

Поступила в редколлегию 01.10.2009

УДК 658.012.011.56

Метод і технологія представлення етимологічного словника української мови у цифровому середовищі / І. В. Остапова // Біоніка інтелекту: наук.-техн. журнал. — 2009. — № 2 (71). — С. 40–47.

Описано технологію побудови комп'ютерного етимологічного словника на основі методології лексикографічних систем. Словник розглядається як інформаційна система, архітектура якої відповідає стандартній архітектурі ANSI/SPARK. На концептуальному рівні використовується лексикографічна модель даних. Розроблений користувацький інтерфейс забезпечує доступ до кожного структурного елемента словника.

Л. 4. Бібл.: 7 найм.

UDK 658.012.011.56

Method and technology of the Ukrainian Etymological Dictionary representations in the digit mood / I. V. Ostapova // Bionics of Intelligence: Sci. Mag. — 2009. — № 2 (71). — P. 40–47.

On the base of methodology of the lexicographical systems the technology for the computer Etymological Dictionary forming is described. The dictionary is considered as the information systems of the standard architecture ANSI/X3/SPARK with the lexicographical model as the conceptual level of data representation. The user interface supports access for even structure element of the dictionary.

Fig. 4. Ref.: 7 items.