

# КРОССЛЕКСИКА: УНИВЕРСУМ СВЯЗЕЙ МЕЖДУ РУССКИМИ СЛОВАМИ

*И.А. Большаков,*

*независимый исследователь, доктор технических наук, профессор,  
почетный профессор Национального политехнического института Мексики*

*E-mail: iabolshakov@gmail.com*

*Адрес: г. Москва, ул. Лесная, 45*

*Создан сверхбольшой русский компьютерный словарь из 292 тысяч слов и выражений и 8 миллионов разнообразных связей между ними. Словарь предельно политематичен, рассчитан на любую аудиторию и запросы на русском и английском языках. Посредством лингвистической и энциклопедической информации он помогает в диалоге редактировать тексты и обучает языку. К словарю также могут обращаться внешние программы.*

**Ключевые слова:** русский язык, компьютерный словарь, связи между словами, политематичность, диалоговый и программный доступ, запросы по-английски.

## 1. Введение

За последнюю четверть века русский язык существенно изменился.

- Пополнилась лексика. Накапливавшиеся в обществе разговорные слова и жаргонизмы выплеснулись на страницы изданий, в рекламу, на телеэкран, в Интернет. Появилось множество новых заимствований, многие слова приобрели новый смысл.

- Соответственно, изменился и пополнился состав словосочетаний, которыми, по формулировке И.Мельчука [1], только и говорит человек.

- Поляризовалась ситуация в части владения языком. На одном полюсе возросло число обозревателей, журналистов и ученых-гуманитариев, вир-

туозно владеющих языком и более не стесненных советскими речевыми штампами. На другом полюсе появилась масса «афтаров», которые демонстрируют в Интернете убогий язык, попирающий все нормы орфографии и приличия.

Речь, конечно, не идет о разрушении языка или «нервном срыве» [2], смежные поколения всегда договорятся друг с другом, но имеющиеся академические словари сильно устарели. Появившиеся крупные словари, например, [3, 4, 5, 6], как-то успевают истолковывать новации, но слабо отражают словосочетания, особенно – новые.

Создание словарей, отражающих допустимые и недопустимые связи между словами, стало крайне актуальным. Даже высоко грамотные люди могут не помнить или не знать, например, что *иск вчиняют*,

а католицизм практикуют. Людей с грамотностью пониже (а таких миллионы) нужно уберечь от «некультурных» выражений, как совсем примитивных (*более лучше, очень прекрасный, играть значение...*), так и более тонких (*поединок команд, более оптимальный, нелицеприятный разговор, возвести баню, раскаться о случившемся, потерпеть несправедливость, болтнуть глупость, принести урон...*). Таких «ляпов» множество и в текстах они нередки. Но есть еще больше языково-специфичных и притом вполне нормативных словосочетаний, которыми плохо владеют как многие с родным русским языком, так и особенно иностранцы: *решение можно принимать и выносить, тревогу испытывать и разделять, внимание обращать и уделять, сочувствие выразить и испытывать*. Чтобы понять русскую специфичность таких выражений, достаточно перевести их на другой язык пословно: правильный перевод обычно не получается. При этом не существует четкой границы между специфичными и неспецифичными (свободными) словосочетаниями. Поэтому желательно подобрать их как можно больше, допустимых и недопустимых. И пусть их число измеряется миллионами и постоянно растет, а цель исчерпывающего подбора всех словосочетаний недостижима.

Возможность отобразить миллионы межсловных связей в едином словаре появилась сравнительно недавно — благодаря быстрому совершенствованию компьютеров. Типовой объем их дискового накопителя увеличился за четверть века в тысячи раз. В памяти десктопа, ноутбука и смартфона умещаются тексты любого нужного объема. При выдаче словарей на экран уже не обязательно повторять их бумажный формат с его обычными сокращениями. Можно активнее использовать цвет, привлекать мультимедиа.

Для нас важно то, что теперь нет необходимости придерживаться привычного *линейного* построения словарей. Согласно этому принципу словарь — это последовательность статей, характеризующих смысл и грамматические категории своего заглавного слова (титула). Связи с другими словами при этом указываются лишь эпизодически. Отстают от этого принципа лишь в особых случаях, например, в тезаурусах и словарях синонимии или антонимии. Повышенный интерес к межсловным связям диктует *комбинаторный* принцип построения словаря, при котором титул, т.е. слово или устойчивое выражение, включается в него вместе со всеми обнаруженными на данный момент его связями с другими титулами.

Настоящая статья описывает новый компьютерный словарь русского языка *КроссЛексика* (CrossLexica), отвечающий возникшим потребностям и воплощающий следующие идеи:

♦ **Комбинаторный** принцип: титул включается только вместе со своими связями.

♦ **Декомпозиционный** принцип: титул — это одиночное слово или словосочетание; в последнем случае полнозначные слова титула являются и отдельными титулами.

♦ **Политематичность**: словарь покрывает лексику большинства сфер использования языка, от высокой теории до низкого быта, содержит как лингвистические, так и энциклопедические знания.

♦ **Всеохватность** целевой *аудитории*: «от генерала до доярки».

♦ **Покрывание трех известных типов связей между титулами — синтаксических** (в словосочетаниях), **семантических** (= смысловых) и **паронимических** (= внешнего сходства).

♦ **Амбивалентность**: информация выдается словарем как пользователю в диалоге, так и внешней программе по ее запросу.

♦ **Встроенность в современный информационный мир**: словарь позволяет составить запрос и послать его в Интернет-поисковик.

♦ **Двуязычность**: есть *англо-русский* подсловарь, позволяющий по-английски запросить данные на русском языке. Можно также видеть английские переводы титулов.

♦ **Языковая многоуровневость**: кроме синтаксических и семантических связей, для каждого титула выдается *морфопарадигма*, т.е. список всех падежных форм для существительного и прилагательного и личных форм для глагола.

На июль 2013 г. КроссЛексика включает 292 тыс. титулов и более 8 млн. связей между ними, что по объему раз в 15 больше словаря Даля и уже совсем не подходит для печати. Словарь правится и пополняется постоянно, и очередная его версия всегда готова к работе с пользователем (*см. рекламу в данном номере*).

## 2. Источники и тематика контента

Источниками создания и пополнения КЛ явились:

♦ Академические словари русского языка, десятки словарей по экономике, бизнесу, электронике,

вычислительной технике, строительству и другим областям.

◆ Поток новостей, политические и научные статьи портала *газета.ру*.

◆ Десятки тысяч справок относительно слов и словосочетаний в Гугле и Яндексe.

◆ Рекламные буклеты, объявления по ремонту и строительству, издания о модах, туризме, персоналиях, автомобилях.

Все указанное накапливалось, структурировалось, размечалось и вводилось в базу словаря с 1990 г., в основном – вручную. Нас интересовала любая тематика, как то:

- Экономика, финансы и бизнес;
- Общественно-политическая сфера: политика, политология, социология, публицистика;
- Техника и технологии: радиоэлектроника, компьютеры, программирование, Интернет, автомобили, авиация, военная техника, бытовая техника, строительство;
- Точные и естественные науки: математика, физика, химия, биология, геология, география, экология;
- Гуманитарные и смежные сферы: психология, философия, история, лингвистика, лингвострановедение, искусство, религия;
- Медицина;
- Спорт;
- Кулинария;
- Бытовой язык, включая обценную лексику без слов трех жестко табуированных русских корней. (Но синонимы и эвфемизмы таких слов Госдума не запрещала!)

В итоге КЛ покрывает нужды самого широкого круга пользователей, включая ученых, преподавателей, инженеров, журналистов, бизнесменов, военных, студентов, школьников, пенсионеров, домохозяйек.

### 3. Титулы, связи между ними, общая структура словаря

Титулы в КЛ относятся к четырем главным частям речи:

#### ■ **Субстантивные:**

- Отдельное существительное: *абазур, битва, бифштекс, благо, блины...*
- Устойчивое именное словосочетание: *алко-*

*гольные напитки, ближнее зарубежье, сельское хозяйство, точка зрения, уровень жизни...*

#### ■ **Глагольные в инфинитивной или личных формах:**

- Одиночный глагол: *говорить, идти, обсуждать, спать, ругать...*
- Глагольный оборот: *навести страх, оказывать внимание, испытать ужас...*

#### ■ **Адъективные**, т.е. прилагательные или причастия:

- Отдельное прилагательное: *абстрактный, абхазский, авансовый, автономный, авантюрный, беж, воздушно-реактивный...*
- Отдельное причастие: *задвинутый, перевезенный, желающий...*
- Адъективный оборот: *бросающийся в глаза, хорошо одетый, бойцовской породы, большой дальности, бывший в употреблении, в елочку, как бархат...*

#### ■ **Адвербиальные**, т.е. наречия или деепричастия:

- Отдельное наречие: *абсолютно, абстрактно, адски, аляповато, быстро...*
- Отдельное деепричастие: *базируясь, надев, торопясь, шепча...*
- Адвербиальный оборот: *аккуратным образом, более или менее, как банный лист, как выжатый лимон, в особой степени, куда попало, мелкой дрожью...*

Предлоги встроены в словосочетания и титулами обычно не являются, а многие прочие служебные слова условно считаются адвербиальными.

Упорядоченную по алфавиту совокупность титулов назовем **словником**.

Связи между титулами бывают трех базовых типов:

■ **Синтаксические** связи формируют словосочетания. Словосочетание у нас – это два полнозначных слова, синтаксически связанных и устойчиво совместимых по смыслу. Связь полнозначных слов может включать служебное слово (предлог или союз) по формуле *полнозначное слово1 → (служебное слово) → полнозначное слово2*, напр., *сотрудничество → ради → мира*, где стрелка означает синтаксическое управление.

■ **Семантические** связи соединяют слова со смысловым сходством. Это синонимы, антонимы и др., см. ниже.

■ **Паронимические** связи соединяют слова либо с буквенным сходством (например, отличающиеся лишь одной буквой слова *кадка* и *каска*), либо с морфемным сходством (например, однокоренные слова *кредитка* и *кредиторка*).

Глобальной структурой КЛ является матрица

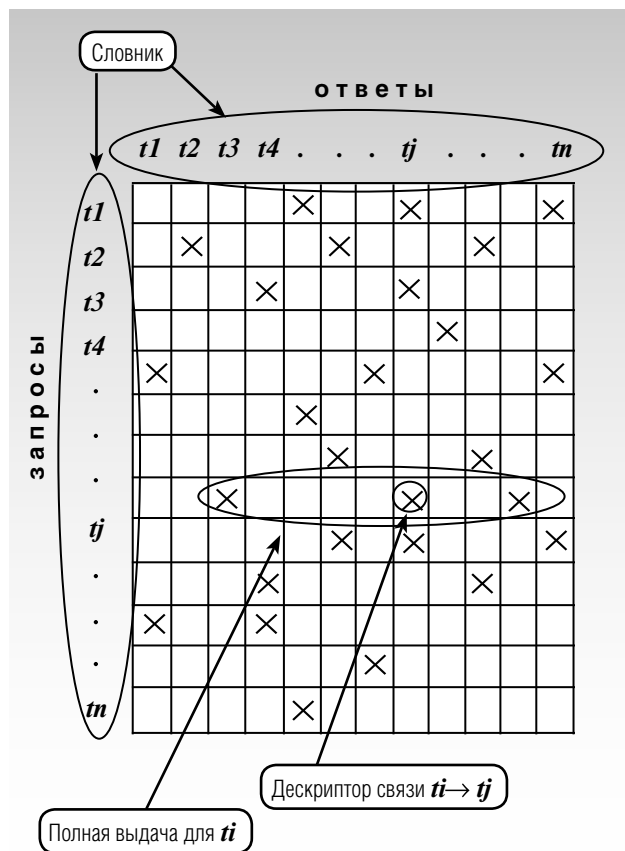


Рис. 1. КроссЛексика как матрица связей

{Словник × Словник} (рис. 1). Слева указаны титулы запросов, сверху – титулы ответов, крестиками отмечены дескрипторы связей между титулами. Связи бывают любого из указанных типов, а их дескрипторы указывают конкретный тип и подтип и включают, где следует, код соединяющего служебного слова. Запрос в виде титула  $t_i$  выхватывает из матрицы титулы ответов  $t_j$ , имеющих связи с  $t_i$ .

На допустимые связи между словами накладывает ограничения как сам язык, так и реалии внешнего мира. В итоге громадная матрица связей из  $8.4 \cdot 10^{10}$  элементов оказывается крайне разреженной: непуст в среднем лишь каждый 10500-й ее элемент. Но это значит, что уже найдено 8 млн. непустых элементов.

#### 4. Словосочетания

Словосочетания составляют самую важную и самую объемную часть КЛ. Их разнообразие сформировано различными подтипами синтаксических связей. Основная масса словосочетаний включает подтипы, зафиксированные в КЛ сотнями тысяч:

◆ Определительная пара «существительное – прилагательное»: *краснокочанная капуста, явный наглец, полная ясность, платье беж...*

◆ Определительная пара «глагол / прилагательное / наречие – наречие»: *полностью ясный, резко высказаться, ужасно страшно...*

◆ «Причастие/прилагательное – его дополнение–существительное»: *рассмотревший вопрос, ковырявший в носу, красный от гнева,...*

◆ «Глагол – его дополнение–существительное»: *ковырять в носу, рассмотреть вопрос, остаться из-за погоды, купить на рынке...*

◆ «Деепричастие/наречие – его дополнение–существительное»: *решив вопрос, ковыряя в носу, близко от города, купив на рынке...*

◆ «Существительное – подчиненное ему существительное»: *сердце матери, отличия в произношении, борьба против терроризма...*

◆ «Подлежащее–существительное – сказуемое в виде личной формы глагола или краткого адъектива»: *самолет вылетел, доклад краток, враг напал, глазки бегают...*

Прочие словосочетания представлены в КЛ десятками или единицами тысяч. Вот наиболее представительные из них:

◆ «Устойчивые сочиненные пары»: *наука и техника, власть и бизнес, ясный и четкий, взвесить и решить, в срок и в полном объеме, авиа- и железнодорожный транспорт...*

◆ «Глагол – его инфинитивное дополнение»: *собраться поехать, хотеть перекусить...*

◆ «Существительное – его инфинитивное дополнение»: *соблазн взять, желание уйти...*

На сегодня накоплено 2.26 млн. словосочетаний, и каждое из них доступно с двух сторон, что эквивалентно 4.52 млн. односторонних связей между словами.

#### 5. Семантические связи

Семантических связей суммарно 2.77 млн. Наиболее широко представлены:

● **Синонимы:** 22.1 тыс. групп по 5.4 элементов, 1.291 млн. связей;

● **Смысловые производные:** 4.2 тыс. групп по 14.6 элементов, 1.096 млн. связей. Вот пример группы смысловых производных: {*извлечение, извлечения; извлекать, извлечь; извлеченный, извлекающий, из-*

влекший; *извлекая, по извлечении, путем извлечения*};

● **Когипонимы** («одноклассники»): 224 тыс. связей. Например, когипонимы для *мясо* – *вырезка, грудинка, гуляш, котлеты, паштет, фарш*;

● **Ассоциации**: 81 тыс. связей. Например, ассоциации для *аденоиды* – *аллергия, бассейн, гомеопатия, кашель, лазеротерапия, слух*.

Прочих семантических связей заметно меньше. Это

○ **Меронимы/холонимы** (= части/целые), напр., *террариум* – *зоопарк*;

○ **Гипонимы/гиперонимы** (= виды/роды), напр., *диплом* – *документ*;

○ **Антонимы**.

Все приведенные типы семантических связей хорошо известны в лингвистике, кроме ассоциаций, являющиеся нашей новацией. Они основаны на сочиненных парах, широко представленных в запросах к Интернет-поисковикам или в их базах данных. Например, ассоциация *аденоиды* – *аллергия* включена, поскольку в поисковиках замечены пары *аденоиды и аллергия, аллергия и аденоиды*. Такие ассоциации полезны для составления запросов к Интернету, на их основе можно строить различные онтологии, а их статистика характеризует профиль массового пользователя Рунета.

Семантические связи в целом полезны следующим:

■ Они помогают понять смысл титула. Среди синонимов много толковательных, например, смысл титула *халляльный* хорошо поясняется его синонимом *отвечающий мусульманским нормам*. Гиперонимы (родовые понятия) давно служат в науке ядрами определений. Увидев, например, для *эндометриоз* гипероним *акушерско-гинекологическая болезнь*, пользователь едва ли захочет полного толкования.

■ Вместе с синтаксическими связями семантические позволяют прогенерировать словосочетания, в словаре отсутствующие. Так, исходя из содержащихся в базе словосочетания *букет цветов* и соотношения **Гипероним** (*каллы*) = *цветы*, для *каллы* порождается словосочетание *букет калл*, в базе отсутствующее. Такие «логические» выводы КЛ совершает миллионами, но из-за негарантированной правильности результаты выводов выдаются на экран низким контрастом.

■ Семантические связи позволяют выдавать разнообразные энциклопедически справки:

○ Названия континентов, океанов, крупных морей, горных цепей и других геообъектов;

○ Названия крупнейших городов мира в привязке к странам;

○ Сведения о 60 странах, по 20 крупнейшим из них – более подробные: столица, валюта, способ правления, титул главы государства, титульная нация, государственный язык, единица административного деления, преобладающая религия;

○ Названия и другие сведения о десятках городов и регионов России, включая наименования жителей: *Тула* – *туляки*, *Архангельск* – *архангелогородцы*, *Курск* – *куряне...*; для Москвы приведены названия всех ее районов;

○ Около 300 наиболее частых русских имен вместе с их диминутивами (*Сергей* – *Серезжа*);

○ Имена известнейших политических, деловых, научных и культурных деятелей мира.

○ Названия наиболее крупных организаций (корпораций) мира.

○ Названия наиболее известных художественных произведений (романов, фильмов, опер, мюзиклов и пр.)

Энциклопедической является также широко представленная терминология экономики, финансов, бизнеса, медицины, точных, естественных и гуманитарных наук.

## 6. Шаги навстречу рядовому пользователю

Для удобства рядового пользователя нужно было упростить интерфейс и сделать ненужным освоение сугубо лингвистических терминов. Для этого было предпринято следующее.

Исключены привычные пометы, используемые в печатных словарях. Оставлено два типа помет – степени разговорности (стиля) и степени фигуральности. Они служат пользователю рекомендациями и побудительными стимулами.

У разговорности пять градаций, размеченных цветом:

<нет пометы> Хорошо бы знать и уметь употребить это слово или выражение (*книга, налоги, конституция, роуминг...*).

<зеленый буллит> Специальное, книжное или забытое слово или выражение (*парадигма, афедрон, квадруольный...*); пользуясь им, опасайтесь непонятности.

<желтый буллит> Чисто разговорное слово или выражение (*мотать нервы, жевать сопли, мочить в сортире...*); не пользуйтесь им в официальных документах. Данная градация получилась слишком обширной, поскольку мы пока не выделяем из нее слова и выражения, имеющие сильную эмоциональную окраску.

<красный буллит> Обсценное слово или выражение (*говно, жопа, мудака, взять за яйца...*); не пользуйтесь им при дамах, детях и в официальной обстановке.

<серый буллит> Бытующее «некультурное» выражение (*оплатить за проезд, пошить пальто...*); правильнее передавать его смысл иначе. (Эквивалентный нормативный вариант, как правило, дается в виде синонима.)

Элементы любой расцвеченной градации можно отбросить в рамках данной сессии.

У фигуральности три градации:

<нет пометы> Понимается как есть (*идти в школу, вызвать слесаря*);

(*fig*) Понимается только фигурально (*сесть в галошу, висеть на волоске*);

(*mb-fig*) Понимается либо фигурально, либо в прямом смысле (*сесть в лужу, первая ракетка*).

Перед началом работы можно выбрать версию интерфейса – научную либо народную. В обеих версиях элементы меню, толкования омонимов и хелп-информация даются по-русски, но разделы выдачи именуются в научной версии строгими терминами (*Синонимы, Гипонимы, Когипонимы...*), а в народной – их «народными» синонимами (*Сходные по смыслу, Подклассы, Одноклассники...*). Реализованы и две аналогичные версии с английскими заголовками.

Малоподготовленному пользователю и иностранцу трудно выделить из множества выдаваемых словосочетания те, что подлежат первоочередному освоению. Но он может выбрать в рамках сессии частотный порядок выдачи определительных конструкций вместо алфавитного. Тогда наиболее многочисленные в КЛ определения предшествуют остальным. Можно вообще отсечь в данной сессии редкие словосочетания.

Запрос можно ввести по-разному: набором с клавиатуры; выбором строки в словнике; выбором строки выдачи на экране (это шаг навигации по словнику); шагом назад–вперед по списку *История*; вводом эквивалентного английского слова.

## 7. Различные приложения словаря

Предусмотрено несколько видов приложений КЛ:

◆ **Диалоговое** (интерактивное) приложение: пользователь вводит запрос в интерактивном режиме и использует выдачу, например, для углубленного изучения русского языка или параллельного редактирования текста за компьютером. Особо ценно для редактирования то, что словарь не только помогает устранять неуместные по стилю обороты, но и дает массу адекватных синонимических перифраз. Поскольку пассивное знание у многих носителей языка значительно превосходит объем активно используемых ими языковых средств, при показе того, как можно выразить ту же мысль иначе, обычно легко обнаруживается более подходящий вариант.

◆ **Интерфейсное** приложение: с помощью КЛ пользователь формирует запрос к Интернету, обращается к нему прямо из словаря и получает результаты поиска.

Оба этих вида приложений реализуются текущей версией КЛ.

В **недиалоговых** приложениях внешняя программа обращается к словарю через специальную утилиту КЛ и использует выдачу самостоятельно. Сами приложения не входят в КЛ и находятся на разных этапах отдельной разработки. В первую очередь эти приложения включают:

◆ **Обнаружение и исправление смысловых ошибок** типа *истерический центр*. В тексте ищутся синтаксически связанные пары слов, не представленные в КЛ и, скорее всего, неосмысленные. Просматриваются паронимы для обоих слов сомнительной пары, и ищутся имеющиеся в КЛ их сочетания. Найденные словосочетания предлагаются пользователю на выбор в качестве исправлений.

◆ **Разрешение неоднозначности омонимов**. Просматриваются словосочетания и семантические связи для омонимичных слов теста, и выбирается тот омоним, для которого в контексте найдено наибольшее число синтаксически и семантически сочетающихся соседей.

◆ **Стеганография и стеганализ**. Текстовые словосочетания и синонимы текстовых слов используются для регулируемой замены одних синонимов другими, дабы этими заменами закодировать стороннюю информацию, которая тем самым тайно передается несущим текстом без изменения смысла носителя и затем декодируется.

◆ **Фильтрация синтаксических разборов**. В ана-

лизируемом предложении отмечаются все словосочетания, наличествующие в КЛ, и чем больше обнаружено таких словосочетаний в данном варианте разбора предложения, тем вероятнее он считается.

Среди других возможных чисто автоматических приложений упомянем определение авторской принадлежности текстов и их общей тематики, рациональное деление текста на абзацы, идиоматичный перевод английских словосочетаний. На базе КЛ можно создавать тематически специализированные «мини-КроссЛексиксы», например, по экономике или вычислительной науке. Можно также строить онтологии, отражающих те многочисленные связи между понятиями современного мира, которые выходят за рамки синонимии, антонимии, гиперонимии / гипонимии и смысловой производности.

### 8. Утилиты создания рабочей версии и интерфейс словаря

При создании КЛ не использовались стандартные СУБД, и матрица связей была воплощена программой на C++. Сейчас КЛ работает под ОС Windows XP и Windows 7.

Отдельно программировались многочисленные утилиты морфоклассификации титулов и кодирования дескрипторов связи на основе исходной текстовой базы. Создана и специальная утилита сборки очередной рабочей версии.

Интерфейс КЛ кардинально отличается от привычных словарей. На рис. 2 представлена типовая выдача и размечены основные составляющие интерфейса: закладки сверху отображают словники, а закладки справа указывают начала различных

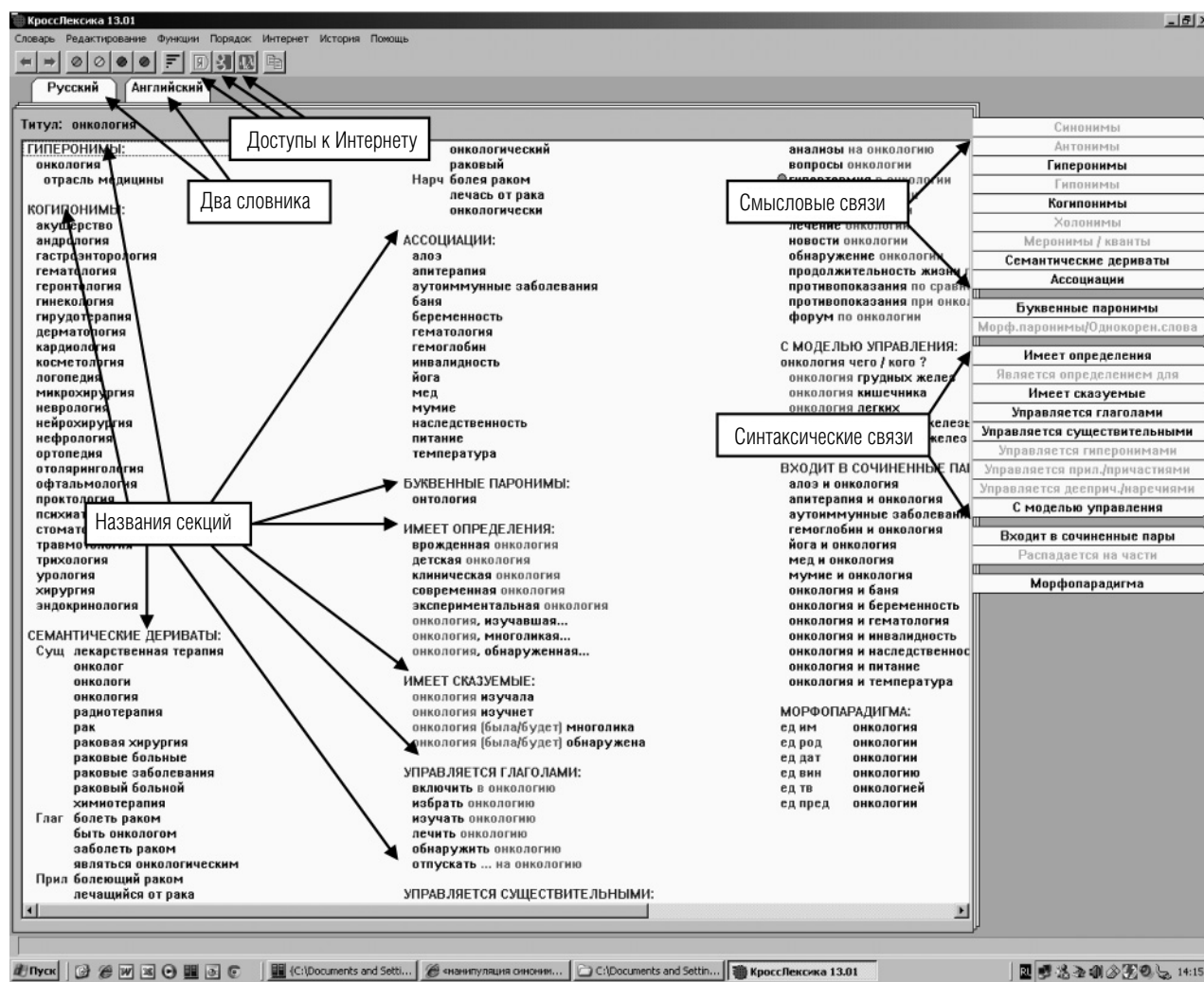


Рис. 2. Выдача для титула «онкология»

секций выдачи, в которых собраны титулы с одинаковым типом дескрипторов связи. В данном примере все секции уместились на одном экране. По верхнему краю экрана идут доступные пользователю функции управления и дублирующие их кнопки.

Если собрать у всех титулов секции синонимов, получим синонимический словарь, по объему превышающий все известные печатные словари. Тем же способом могут быть получены крупные словари антонимов и семантических дериватов, тезаурусы гипонимов / гиперонимов (= надклассов / подклассов) и холонимов / меронимов (= частей / целых), словари глагольно-именного управления, сочиненных пар и др. Тем самым, КЛ объединяет всех их под одной крышей.

КЛ не содержит толкований всех титулов подряд. Есть лишь краткие толкования омонимов, многочисленные семантические связи и английские переводы титулов. За толкованиями в процессе диалога можно обращаться через КЛ, например, в Викисловарь, но по объему обработанного словника он пока существенно уступает КЛ. В недиалоговых приложениях КЛ вполне самодостаточна, поскольку обычные словарные толкования, пусть даже оцифрованные, для современных автоматических средств обработки текстов бесполезны.

## 9. Заключение

Создан новый компьютерный словарь – ресурс, по объему и структуре не имеющий аналогов ни для одного языка мира. В то время как в [7] собрано 270 тыс. русских словосочетаний для 2.5 тыс. слов, а в [8] – 250 тыс. английских словосочетаний для 9 тыс. слов, словник КроссЛексики на порядок больше, словосочетаний здесь 2.26 млн., отражены также миллионы несинтаксических связей, и все это выдается в рамках единого компьютерного интерфейса.

КроссЛексика уже готова как вспомогательное средство редактирования сложных текстов любой тематики и обучения русскому языку в его деловой, литературной и разговорной ипостасях. Также КЛ позволяет поднять существующие типы систем автоматической обработки русских текстов на уровень, ныне недоступный.

Применительно к грядущему словарю [9], реализующему заметно более глубокий, подлинно академический анализ связей между русскими словами, КроссЛексику можно рассматривать как обширнейший массив сырого материала, который, правда, уже отвечает нуждам широкого круга пользователей-лингвистов.

Выражаю благодарность А.Ф.Гельбуху, кто спроектировал и запрограммировал интерфейс словаря и много лет бескорыстно помогал мне в многотрудном деле. ■

## Литература

1. Mel'čuk I. Phrasemes in Language and Phraseology in Linguistics // M.Everaert et al. (Eds.) Idioms: Structural and Psychological Perspectives. – Lawrence Erlbaum Associates Publ.: NJ / Hove, 1995. – P. 169-252.
2. Кронгауз М.А. Русский язык на грани нервного срыва. – М.: Языки славянской культуры. Знак, 2008.
3. Толковый словарь русского языка начала XXI века. Под ред. Г. Н. Складчиковой. – М.: Эксмо, 2007.
4. Крысин Л.П. Толковый словарь иноязычных слов. – М.: Эксмо, 2008.
5. Большой толковый словарь русских существительных. Под ред. Л.Г.Бабенко. – М.: АСТ-Пресс Книга, 2009.
6. Толковые словари LINGVO, компания ABBY, 2012.
7. Словарь сочетаемости слов русского языка. Под ред. П.Н.Денисова и В.В.Морковкина. – М.: Русский язык, 1983.
8. Oxford Collocation Dictionary for students of English. – Oxford, NewYork: Oxford University Press, 2009.
9. Апресян В.Ю., Апресян Ю.Д., Бабаева Е.Э., Богуславская О.Ю., Галактионова И.В., Гловинская М.Я., Иомдин Б.Л., Крылова Т.В., Левонтина И.Б., Птенцова А.В., Санников А.В., Урысон Е.В. Проспект активного словаря русского языка. Под ред. акад. Ю.Д.Апресяна. – М.: Языки славянских культур, 2010.