

Кушубек кызы Бегимай, Карабаева С.Ж., Сагынбекова Гульназ

И.Арабаев атындагы КМУнун маалыматтык технологиялар кафедрасынын магистранты,
ф.и.к, Н.Исанов атындагы КМКТАУ доценти

И.Арабаев атындагы КМУнун маалыматтык технологиялар кафедрасынын магистранты

Кушубек кызы Бегимай, Карабаева С.Ж., к.ф.н, доц., Сагынбекова Гульназ

Магистрант кафедрасы информатсионных технологий КГУ им. И. Арабаева,
К.ф.н., доцент КГУСТА им. Н. Исанова,

Магистрант кафедрасы информатсионных технологий КГУ им. И. Арабаева

Kushubek kyzy Begimai, Karabaeva S.Zh.,Sagynbekova Gulnaz

Master student of the Department of Information Technologies of KSU named after I. Arabaeva,
Ph.D., Associate Professor of KSUST named after N. Isanova,

Master student of the Department of Information Technologies of KSU named after I. Arabaeva

ЛЕКСИКОГРАФ ЛЕКСИКАЛЫК ЖЕЛЕ ТОКУП ЖАТКАН СЫЯКТУУ

ЛЕКСИКОГРАФ КАК ПЛЕТУЩИЙ СВОЮ ЛЕКСИЧЕСКУЮ СЕТЬ

LIKE A LEXICOGRAPHER WEAVING HER LEXICAL NETWORK

Аннотация: Макалада лексикалык жеткиликтүүлүктүн когнитивдик ыкмалары, сөздөрдүн мааниси жана онтология жөнүндө ойлор, лексикондорду кол жана автоматтык түзүүдөгү ыкмалар, ошондой эле прагматикалык ыкмалар жана аспекти талкууланат. Лексикалык ресурстар үчүн жаңы маалымат структуралары, лексикалык моделдерди түзүүнүн жаңы ыкмалары менен айкалыштырылышы керек.

Аннотация: В статье рассмотрены когнитивные подходы к лексическому доступу, соображения по поводу значения слов и онтологий, ручные и автоматические подходы к построению лексиконов, а также прагматические подходы и аспекты. Новые структуры данных для лексических ресурсов должны сочетаться с новыми способами построения лексических моделей.

Annotation: The article discusses cognitive approaches to lexical access, considerations about the meaning of words and ontologies, manual and automatic approaches to building lexicons, as well as pragmatic approaches and aspects. New data structures for lexical resources should come together with new ways of building lexical models.

Негизги сөздөр: лексика, лексикография, лексикограф, лексикалык тармак, өрмөк тору, лексикалык ресурс, нейрон моделдери, нейросемантика, нейрокомпьютердик теория, кластерлер, социалдык тармак, семантикалык издөө, контексте байланыштуу издөө, интерфейс, маалыматтарды визуалдаштыруу, жарым автоматтык киргизүү, катыштык маалыматтар базалары.

Ключевые слова: лексика, лексикография, лексикограф, лексическая сеть, паутина, лексический ресурс, нейронные модели, нейросемантика, нейровычислительная теория, кластеризация, социальная сеть, семантический поиск, контекстно-зависимый поиск, интерфейс, визуализация данных, полуавтоматический ввод, реляционные базы данных.

Key words: vocabulary, lexicography, lexicographer, lexical network, web, lexical resource, neural models, neurosemantics, neurocomputational theory, clustering, social network,

semantic search, context-sensitive search, interface, data visualization, semi-automatic input, relational databases.

За последнее десятилетие NLP (Natural Language Processing) стало свидетелем всплеска развития многоязычных лексических баз данных и инструментов для кросс-языковых задач, таких как поиск информации, машинный перевод и овладение иностранными языками.

Название нашей беседы - неявная ссылка на английское клише, как паук, плетущее свою паутину, - призвано привлечь внимание к метафоре, которая может быть проведена между танцем паука, плетущего свою паутину, и новым лексикографическим жестом, который постепенно появляется. из работ по лексическим ресурсам типа WordNet (Fellbaum, 1998; Baker et al., 2003; Gader et al., 2012).

Однако словарь - будь то в бумажном или электронном формате - является лишь одним из многих возможных воплощений лексических моделей.

Мы утверждаем, что графическая структура, присущая лексиконам естественного языка, определяет не только приобретение и использование словарного запаса, но и лексикографическую активность. В этом отношении размышления о новых способах реализации задачи построения лексических ресурсов важны как для самих лексикографов, так и для всех, кто интересуется лексиконами как ментальными структурами. В конце концов, лексикографы и изучающие язык - это те, кто имеет самый непосредственный контакт с лексическими структурами посредством тесно связанных действий: описание природного явления - это форма обучения посредством явной концептуализации. Лексикографы часто сталкиваются с тем, что, завершая описание слова, они достигают определенной формы понимания и усвоения этого слова. Они не просто транскрибируют знания слов и наблюдения, сделанные над поведением слов в речи и текстах: они «усваивают» слово. Это заставляет их чувствовать себя хорошо, и это объясняет, почему лексикография действительно вызывает сильное привыкание.

Паутину можно рассматривать как ловушку, в которую попадаешь. Это, так сказать, точка зрения мухи или невинного жука. Однако лексикографу не следует попадаться в сеть: они могут вести себя как пауки, плетущие паутину. Это возможно, если модель, которую они строят, действительно является схематическим представлением - в семиотическом смысле (Farias and Queiroz, 2006) - изучаемой лексики естественного языка. Когда лексикографы бегают по страницам, пишут словарные статьи, как мухи, идущие по стеклянному окну, у них больше всего шансов попасться в паутину слов. Вот почему лексикографы давно ввели системы карточек и записей, чтобы помочь им собрать данные о лексических единицах. Лексикографические записи помогли лексикографам вырваться из двумерной тюрьмы словаря. Их знания о словах занимали «объем» картотеки, что больше соответствовало трехмерной природе лексиконов, которые им приходилось описывать. Позже, с появлением вычислительной лексикографии, реляционные базы данных заменили картотеки в качестве удобных инструментов . . . и метафоры.

Согласно общепринятому мнению, лексикография - это все о написании словарей, и, следовательно, любая деятельность, направленная на построение других типов лексических моделей, освобожденных от двумерного (текстового) словаря, не является «истинной» лексикографией.

Это заблуждение, очень распространенное среди непрофессионалов и поддерживаемое многими исследователями естественного языка, происходит главным

образом из-за того факта, что на протяжении веков у лексикографов не было лучшего средства кодирования, чем текст, и не было лучшей физической поддержки для их описания, чем скрепленные вместе листы бумаги (словари).

Что действительно необходимо и достаточно для того, чтобы задача называлась лексикографией, так это того, что:

- он нацелен на описание лексических единиц одного или нескольких естественных языков с точки зрения смысла, форм и всех других соответствующих языковых свойств;
- в нем используется четко определенная структура анализа, позволяющая согласованно и единообразно описывать все лексические единицы;
- это, по сути, ручная задача, но без ограничений по количеству и разнообразию инструментов и внешних данных, которые можно использовать для выполнения этой задачи;
- он «видит большое»: чем больше охват и глубина описания каждой лексической единицы, тем более лексикографической будет задача.

Последний пункт более важен, чем может показаться: когда дело касается лексики - ее описания, а также изучения, усвоения и т. д. - размер имеет значение. Возьмем крайний случай: человека, чей единственный опыт в данной области - это описание только одной или нескольких лексических единиц, нельзя считать лексикографом, а выполненная задача - это всего лишь упражнение в лексикографии. Напротив, тот, кто достиг описания десятков тысяч лексических единиц, несомненно, является опытным лексикографом. Где-то посередине происходит переход от ученика к настоящему лексикографу.

Обратите внимание, что в приведенной выше характеристике лексикографии не упоминается формальная природа лексических моделей. Фактически, когда инициаторами WordNet было предложено создание совершенно новой графической модели лексического знания (Miller et al., 1990), не было сделано никаких заявлений о появлении новой дисциплины. Напротив, лексикография оставалась эталоном, а работу выполняли люди, называемые лексикографами, которые создавали наборы данных, называемые файлами лексикографов. И это вполне оправданно, ведь именно лексикография - это не словари как таковые.

Словари хранят знания о словах. Очевидно, что они должны быть исчерпывающими и достаточно полными, чтобы раскрывать значение слов (анализ), их форму или другую связанную информацию, актуальную для производителей языка (говорящих, писателей). Однако качество словаря зависит не только от охвата, но и от доступности информации. Стратегии доступа различаются в зависимости от задачи (понимание текста или создание текста) и знаний, доступных на момент консультации (слова, концепции, звуки речи). В отличие от читателей, ищущих смыслы, писатели начинают с них, ища соответствующие слова. В то время как бумажные словари статичны и допускают только ограниченные стратегии доступа к информации, их электронные аналоги обещают динамический, упреждающий поиск по множеству критериев (значение, звук, родственные слова) и через различные маршруты доступа. Навигация происходит в огромном концептуальном лексическом пространстве, и результаты отображаются во множестве форм (например, в виде деревьев, списков, графиков или отсортированы в алфавитном порядке, по теме, по частоте).

Наш взгляд на словари (их создание и использование) кардинально изменился за последние 30 лет. В прошлом они рассматривались как приложение к грамматике, но тем временем переместились в центр внимания. В самом деле, в NLP вряд ли можно выполнить какую-либо задачу без них. Кроме того, вместо того, чтобы быть статическими объектами

(представление базы данных), словари теперь рассматриваются как графы, узлы и связи которых (сила соединения) могут изменяться со временем. Интересно, что свойства топологии, кластеризации и эволюции, известные из других дисциплин (общество, экономика, человеческий мозг), также применимы к словарям: все связано, следовательно, доступно, и все развивается.

Анализ концептуального ввода пользователя словаря

- С чего начинается языковой продюсер (набор слов)?
- Что думают авторы, когда создают сообщение и ищут слово?
- Что нужно для преодоления разрыва между этими вводными и желаемыми результатами (целевой слово)?

Значение слов

- Лексическое представление (целостное, разложенное)
- Представление смысла (на основе понятий, примитивы)
- Выявление скрытой информации (векторные подходы)
- Нейронные модели, нейросемантика, нейровычислительные теории представления содержания.

Структура лексики

- Обнаружение структур в лексиконе: формальная и семантическая точки зрения (кластеризация, тематическая структура)
- Творческие способы получения доступа к словесным ассоциациям и их использования.
- Эволюция, т.е. динамические аспекты лексики (изменение весов)
- Нейронные модели ментального лексикона (распределение информации о словах, организация слов)

Способы создания словарей или указателей

- Ручное, автоматическое или совместное создание словарей и индексов (семантика распределения, краудсорсинг, серьезные игры и т. Д.)
- Влияние и использование социальных сетей (Facebook, Twitter) для создания словарей, для организации и индексации данных (кластеризация слов), а также для отслеживания навигационных стратегий и т. Д.
- (Полу) автоматический ввод типа ссылки (например, синоним, ассоциация, словосочетание, ...)
- Использование корпусов и шаблонов (интеллектуальный анализ данных) для получения доступа к словам, их использованию, комбинациям и ассоциациям.

Доступ к словарю (стратегии навигации и поиска), проблемы с интерфейсом

- Семантический поиск
- Поиск (простой запрос против нескольких слов)
- Контекстно-зависимый поиск (изменение целей пользователей во время поиска)
- Восстановление
- Навигация (часто используемые шаблоны навигации или поисковые стратегии)
- Проблемы интерфейса, визуализация данных.

Мы предлагаем вышеприведенные наблюдения как данность, включая тот факт, что лексикография действительно должна быть нацелена на виртуальные словари, созданные из нетекстовых лексических моделей.

Список использованной литературы:

1. Pankov P.S., Karabaeva S.J. Mathematical and computer models of spatial concepts in Kyrgyz language // Интернет-журнал ВАККР, 2016, № 3. – 7 с.
2. Karabaeva S.J., Pankov P.S. Independent computer presentation of spatial notions in Turkic languages // Пятая Международная конференция по компьютерной обработке тюркских языков «TurkLang 2017». – Труды конференции. Том 1. – Казань: Издательство АН Республики Татарстан, 2017. - С. 68-78.
3. Карабаева С. Единый алгоритм словоизменения и представление пространства в кыргызском языке. – Saarbrücken, Deutschland: Lap Lambert Academic Publishing, 2016. – 62 с.
4. Карабаева С. Ж. Виртуальные геометрические объекты, создаваемые глаголами в кыргызском языке // В мире науки и искусства: вопросы филологии, искусствоведения и культурологии: сб. статей по материалам LXVI междунар. научно-практ. конференции, № 11(66). – Новосибирск: СибАК, 2016. – С. 74-79.
5. Zadeh L. A. The concept of a linguistic variable and its application to approximate reasoning // Information Sciences, 1975, Vol. 8, pp. 199-249, 301-357; Vol. 9, pp. 43-80.
6. Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., and Miller, K. J. (1990). Introduction to WordNet: An On-line Lexical Database. International Journal of Lexicography, 3(4):235–244.
7. Farias, P. and Queiroz, J. (2006). Images, diagrams, and metaphors: Hypoicons in the context of Peirce's sixty-six-fold classification of signs. Semiotica, 162(1/4):287–307.

Рецензент: к.филол.н., доцент Жумалиева Г.Э.