



ҚАЗАҚСТАН РЕСПУБЛИКАСЫ
ТҰҢҒЫШ ПРЕЗИДЕНТІ - ЕЛБАСЫНЫҢ ҚОРЫ

«ҒЫЛЫМ ЖӘНЕ БІЛІМ – 2017»

студенттер мен жас ғалымдардың
XII Халықаралық ғылыми конференциясының
БАЯНДАМАЛАР ЖИНАҒЫ

СБОРНИК МАТЕРИАЛОВ

XII Международной научной конференции
студентов и молодых ученых
«НАУКА И ОБРАЗОВАНИЕ – 2017»

PROCEEDINGS

of the XII International Scientific Conference
for students and young scholars
«SCIENCE AND EDUCATION - 2017»



14th April 2017, Astana



**ҚАЗАҚСТАН РЕСПУБЛИКАСЫ БІЛІМ ЖӘНЕ ҒЫЛЫМ МИНИСТРЛІГІ
Л.Н. ГУМИЛЕВ АТЫНДАҒЫ ЕУРАЗИЯ ҰЛТТЫҚ УНИВЕРСИТЕТІ**

**«Ғылым және білім - 2017»
студенттер мен жас ғалымдардың
XII Халықаралық ғылыми конференциясының
БАЯНДАМАЛАР ЖИНАҒЫ**

**СБОРНИК МАТЕРИАЛОВ
XII Международной научной конференции
студентов и молодых ученых
«Наука и образование - 2017»**

**PROCEEDINGS
of the XII International Scientific Conference
for students and young scholars
«Science and education - 2017»**

2017 жыл 14 сәуір

Астана

УДК 378

ББК 74.58

Ғ 96

Ғ 96

«Ғылым және білім – 2017» студенттер мен жас ғалымдардың XII Халықаралық ғылыми конференциясы = The XII International Scientific Conference for students and young scholars «Science and education - 2017» = XII Международная научная конференция студентов и молодых ученых «Наука и образование - 2017». – Астана: <http://www.eni.kz/ru/nauka/nauka-i-obrazovanie/>, 2017. – 7466 стр. (қазақша, орысша, ағылшынша).

ISBN 978-9965-31-827-6

Жинаққа студенттердің, магистранттардың, докторанттардың және жас ғалымдардың жаратылыстану-техникалық және гуманитарлық ғылымдардың өзекті мәселелері бойынша баяндамалары енгізілген.

The proceedings are the papers of students, undergraduates, doctoral students and young researchers on topical issues of natural and technical sciences and humanities.

В сборник вошли доклады студентов, магистрантов, докторантов и молодых ученых по актуальным вопросам естественно-технических и гуманитарных наук.

УДК 378

ББК 74.58

ISBN 978-9965-31-827-6

©Л.Н. Гумилев атындағы Еуразия
ұлттық университеті, 2017

dependence-is-illness-of-modern-society-7301619.html/

6. Nathaniel Mott. (2013, March 12). Evgeny Morozov and Nicholas Carr on how we're all addicted to the Web. Pando. Retrieved from <http://pando.com/2013/03/12/evgeny-morozov-and-nicholas-carr-on-how-were-all-addicted-to-thw-web/>
7. Athanasius Koliofutis. (2013, July 17). Internet addiction of adolescents. Pemptousia. Retrieved from <http://www.pemptousia.ru/2013/07/интернет-зависимость-подростков-ста/>
8. Malygin. (2008, n.p.). Reasons of internet addiction. Psychotherapy center Dr. Malygin. Retrieved from <http://www.psypodderjka.ru/content/view/84/>
9. Get out of the internet immediately. (2014, July 2). Lenta.ru. Retrieved from <http://lenta.ru/photo/2014/07/02/chinacomputers/#0>

УДК 004.912 + 811.512.122-322

ПРОБЛЕМЫ ПОСТРОЕНИЯ ФУНКЦИОНАЛЬНОЙ МОДЕЛИ ТЮРКСКИХ ЯЗЫКОВ

Абдалиев Бауыржан Жумашулы, Еримбетова Айгерим Сембековна

Abdaliev.bz@gmail.com, aigerian@mail.ru

магистрант специальности 6М070300 Информационные системы ЕНУ
им.Л.Н.Гумилева, Астана, Казахстан; аспирант НГУ, Новосибирск, Россия
Научный руководитель – Сагнаева С.К.

Проблема поиска информации на естественном языке (ЕЯ), релевантной информационным потребностям пользователя, остается актуальной в настоящее время в связи с экспоненциальным ростом информационных ресурсов. Актуальной также остается задача создания средств полнотекстового поиска, автоматической классификации и реферирования текста, автоматизированного машинного перевода.

При построении функциональной модели ЕЯ, стадии его лингвистического анализа включают графематический, морфологический, синтаксический и семантический анализ. В результате формируются модели текста, адекватно отражающие его грамматические и смысловые конструкции. Наиболее успешные результаты получены в области морфологического анализа текстов, которые применяются в поисковых машинах Интернета, текстовых редакторах и т.д. Задачи синтаксического и семантического анализа решены не в полной мере. Рассмотрим более подробно первую задачу.

Целью синтаксического анализа является построение синтаксической структуры каждого предложения входного текста путем сопоставления линейной последовательности лексем (слов, фраз) языка с его формальной грамматикой. Грамматика и словарь задают лингвистическую модель языка. Разделение алгоритмов, т.е. механизмов оперирования с имеющимися словарями и грамматиками, от самих грамматик важно тем, что такой подход позволяет менять правила грамматики, не изменяя алгоритмов\программ, работающих с грамматиками. Примерами таких систем являются системы, работающие с контекстно-свободными (КС) грамматиками, в которых модель языка - грамматика с конечным числом состояний, а алгоритм обеспечивает для произвольно взятого предложения построение дерева вывода по правилам грамматики. Если выводов несколько, то перечисляются все. Такой типа алгоритм, представляющий формальную систему, называется анализатором.

Одним из возможных подходов является использование представления структуры текста на основе диаграмм синтаксического анализатора Link Grammar Parser (LGP). Отметим, что в данный момент исследования в значительной мере ориентированы на англоязычные источники.

LGP – это синтаксический анализатор английского языка, разработанный в университете Карнеги-Меллона. Его основой является некоторая неклассическая теория синтаксиса. Получив предложение, система приписывает к нему синтаксическую структуру

(диаграмму), которая состоит из множества помеченных связей (коннекторов), соединяющих пары слов. Имеется вариант системы Link Grammar Parser для английского, русского, немецкого и ряда других языков. В научной литературе имеется также информация о разработках для тюркских языков, казахского и турецкого. В работе У.А. Тукуева и др. [1] рассматривается система машинного перевода с казахского на английский язык и наоборот с использованием грамматики связей и статистического подхода. Грамматика связей играет важную роль в предложенном алгоритме. Статистический подход применяется для перевода многозначных слов. Разработанные модели и алгоритмы были реализованы в программе машинного перевода. В [1] приведен список из 13 связей, которые естественным образом отражают наиболее важные синтаксические отношения между словами в предложениях на казахском языке. Важным является то обстоятельство, что эти же связи могут быть использованы при разработке синтаксических анализаторов для других тюркских языков, ввиду большой близости не только их синтаксиса, но также морфологии и лексики.

В работе [2] описывается «статистический парсер» зависимостей турецкого языка, базирующийся на использовании статистических моделей обучения на предложениях турецкого языка из так называемого Turkish Dependency Treebank. В результате работы парсер выдает отношения зависимости между лексическими единицами внутри подмножеств множества слов предложения, называемых флективными группами. То есть в отличие от системы Link Grammar Parser, в которой используются словари, содержащие спецификации описывающие связи, в данном случае грамматика связей является производной от статистики.

В [3] рассматривается синтаксический анализатор связей турецкого языка, который, в действительности, является «не лексическим анализатором». Сначала применяется морфологический анализатор, и исходным словам сопоставляются некоторые морфологические описания. Эти описания строятся на основе анализа суффиксов слов, что для агглюнативных языков является естественным. Сохраняются лексемы только отдельных важных в функциональном отношении слов. Далее устанавливаются связи между морфологическими описаниями, а не между исходными словами. По-видимому можно вернуться к исходному предложению и перенести полученные связи на слова, но в работе это не рассматривается. Предложенный в работе подход используется для описания турецкой грамматики в терминах связей, но очевидно, что он применим и к другим тюркским языкам.

Диаграммы, полученные анализатором Link Grammar Parser, представляют из себя графы. Далее следует их предварительная подготовка к сравнению и непосредственно само сравнение. При этом сравнение графов осуществляется не только на лексическом уровне, но и на уровне связей между словами. При сравнении связей используется механизм приоритетов.

Далее предлагаются методы, которые обобщают подход, используемый в базовом алгоритме отождествления предложений. Более точно, базовый алгоритм учитывает только так называемые инвариантные коннекторы, не принимая во внимание более сложную логику. В данном же случае применяются более тонкие методы. При сопоставлении двух предложений, точнее, при анализе их на близость осуществляется проверка ряда логических свойств. Примеры такого рода свойств: инвариантность коннектора, замена коннектора на дизъюнкцию других, расщепление коннектора на два коннектора, расщепление коннектора на два коннектора с инверсией.

Из достоинств системы LGP нужно отметить, что организация самой процедуры нахождения вариантов синтаксического представления очень эффективна. Построение идет не сверху вниз (top-down) и не снизу вверх (bottom-up), а все гипотезы отношений рассматриваются параллельно: сначала строятся все возможные связи по словарным формулам, а потом выделяются возможные подмножества этих связей. [4]

Это, конечно, приводит к алгоритмической непрозрачности системы, поскольку очень трудно проследить за всеми отношениями сразу. Во-вторых, не к линейной зависимости

скорости алгоритма от количества слов, а к экспоненциальной, поскольку множество всех вариантов синтаксических структур на предложении из N слов в худшем случае равномощно множеству всех остовных деревьев полного графа с N вершинами.

Последняя особенность алгоритма заставляет разработчиков использовать таймер, для того чтобы вовремя останавливать процедуру, которая работает слишком долго. Однако все эти недостатки с лихвой компенсируются лингвистической прозрачностью системы, в которой с одинаковой легкостью прописываются валентности слова, причем порядок сбора валентностей внутри алгоритма принципиально не задается – связи строятся как бы параллельно, что полностью соответствует нашей языковой интуиции.

Отметим также отрицательные моменты.

1. Практическое тестирование системы показывает, что при анализе сложных предложений, длина которых превышает 25-30 слов, возможен комбинаторный взрыв, и результатом работы анализатора становится «панический» граф, как правило, случайный вариант синтаксической структуры, с лингвистической точки зрения неадекватной.

2. Применение описанных выше идей затруднено для флективных языков типа русского, ввиду значительно возрастающего объема словарей, которые возникают в силу морфологической развитости флективных языков. Каждая морфологическая форма должна описываться отдельной формулой, где нижний индекс входящего в нее коннектора должен будет обеспечивать процедуру согласования. Это приводит к усложнению набора коннекторов и к увеличению их количества. Для агглюнативных языков (например, тюркских) система станет еще более сложной.

Многие исследователи, например [5,6], склоняются к необходимости проведения глубокого семантического анализа текстов для создания их семантических образов, на основе которых можно проводить тонкое ранжирование документов при организации поиска в тексте. Этот подход, несомненно, наиболее разумный, однако требует тщательной и долгой работы над созданием соответствующих теорий и подходящих инструментов для автоматической обработки текстов [7]. В частности, может потребоваться детальное описание различных областей знаний. Поэтому имеет смысл также поиск частичных решений.

Основная задача состоит в построении алгоритмов, которые, проникая в структуру текста, могут вывести адекватную оценку релевантности текста поисковому запросу. Важно, чтобы данная оценка была основана на контексте поискового запроса и не ограничивалась только ключевыми словами, их близостью или частотой.

Для решения такого рода задач можно использовать семантико-синтаксические отношения между словами предложения, получаемые на выходе программной системы Link Grammar Parser [8, 9]. В работах [10,11,12] описан способ (базовый алгоритм) вычисления степени совпадения естественно-языковых конструкций. Исследования полностью были ориентированы на англоязычные источники. На основе вышеупомянутых идей ранее была реализована информационно-поисковая система (ИПС) iNetSearch для английского языка. Проведенное тестирование системы iNetSearch показало эффективность предложенного алгоритма в решении задач поиска информации. Далее были предложены методы, которые обобщают подход, используемый в базовом алгоритме, и позволяющие учесть перефразирования. В настоящее время прилагаются усилия, чтобы развить идеи применительно к тюркским языкам.

Список использованных источников:

1. Tukeyev U.A., Melby A.K., Zhumanov Zh.M. Models and algorithms of translation of the Kazakh language sentences into English language with use of link grammar and the statistical approach // Proc. of IV Congress of the Turkic World Math. Society, 1-3 July, Baku, 2011. – P.474.
2. Eryiğit G., Oflazer K. Statistical Dependency Parsing of Turkish // Proc. of EACL 2006, 11th Conf. of the European Chapter of the Association for Comp. Linguistics, Trento, Italy, 2006. – P. 89-96

3. Istek O, Cicekli I. A Link Grammar for an Agglutinative Language // Proc. of Recent Advances in Natural Language Processing (RANLP 2007), Borovets, Bulgaria, 2007. – P. 285-290
4. Мурзин Ф.А., Батура Т.В., Семич Д.Ф., Сагнаева С.К., Еримбетова А.С., Бакиева А.М., Митьковская М.В., Семенова Н. А. Қазақ және түрік тілдері негізінде байланыстар грамматикасын зерттеу // Вестник КазНУ, – №4 (116), – 2016, – С. 684-691
5. Salton G. Automatic Information Organization and Retrieval, 1968, 514 p.
6. Лезин Г.В., Тузов В.А. Семантический анализ текста на русском языке: семантико-синтаксическая модель предложения // Экономико-математические исследования: математические модели и информационные технологии. – СПб.: Наука, 2003. – Вып. 3. – С. 282-303.
7. Батура Т.В., Мурзин Ф.А. Машинно-ориентированные логические методы отображения семантики текста на естественном языке: моногр. / Институт систем информатики им. А.П. Ершова СО РАН. Новосибирск: Изд. НГТУ, 2008. – 248 с.
8. Temperley D., Sleator D., Lafferty J. Link Grammar Documentation [Electronic resource]. – 1998. – Mode of access: <http://www.link.cs.cmu.edu/link/dict/index.html> (accessed 15 November 2012)
9. Sleator D., Temperley D. Parsing English with a Link Grammar. Pittsburgh: School of Computer Science Carnegie Mellon University, 1991. – 93 p.
10. Murzin F., Perfliev A., Shmanina T. Methods of syntactic analysis and comparison of constructions of a natural language oriented to use in search systems // Bull. Nov. Comp. Center, Comp. Science, 2010, Iss. 31, – P. 91-109.
11. Перфильев А.А., Мурзин Ф.А., Шманина Т.В. Методы синтаксического анализа и сопоставления конструкций естественного языка, ориентированные на применение в информационно-поисковых системах // Вестник НГУ, том 9, выпуск 4, 2011. – С 50-59.
12. Батура Т.В., Мурзин Ф.А., Перфильев А.А., Шманина Т.В. Методы повышения эффективности поиска информации на основе синтаксического анализа // Моногр. / Институт систем информатики им. А.П. Ершова СО РАН. – Новосибирск: Изд-во СО РАН, 2014. – 76 с.

УДК 004.853

ОНТОЛОГИЧЕСКАЯ МОДЕЛЬ ИНТЕЛЛЕКТУАЛЬНОГО ЭЛЕКТРОННОГО УЧЕБНИКА ПО ДИСЦИПЛИНЕ «ПРОГРАММНАЯ ИНЖЕНЕРИЯ»

Абжанова А. А.

Евразийский национальный университет имени Л.Н.Гумилева, Астана, Казахстан
Научный руководитель – Ниязова Р.С.

Важнейшим инфраструктурным компонентом информатизации высшего образования является разработка электронных средств поддержки процесса обучения. В этом широком спектре программных продуктов наиболее востребованными на сегодняшний день являются электронные учебные издания (ЭУИ) [1]

Электронными учебными изданиями называют такие издания, которые предназначены для автоматизации обучения и контроля знаний, соответствующие учебному курсу или отдельным ее частям, а также позволяющие определить траекторию обучения и обеспечивающие различные виды учебных работ. Они подразделяются на сервисные программные средства общего назначения; программные средства для контроля и измерения уровня знаний, умений и навыков обучающихся; электронные тренажеры; программные средства для математического и имитационного моделирования; программные средства лабораторий удаленного доступа и виртуальных лабораторий; информационно-поисковые справочные системы; автоматизированные обучающие системы (АОС); электронные учебники (ЭУ) и электронные учебные пособия; экспертные и интеллектуальные обучающие