

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



Студенттер мен жас ғалымдардың
«ҒЫЛЫМ ЖӘНЕ БІЛІМ - 2016» атты
XI Халықаралық ғылыми конференциясының
БАЯНДАМАЛАР ЖИНАҒЫ

СБОРНИК МАТЕРИАЛОВ
XI Международной научной конференции
студентов и молодых ученых
«НАУКА И ОБРАЗОВАНИЕ - 2016»

PROCEEDINGS
of the XI International Scientific Conference
for students and young scholars
«SCIENCE AND EDUCATION - 2016»

2016 жыл 14 сәуір
Астана

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

**Студенттер мен жас ғалымдардың
«Ғылым және білім - 2016»
атты XI Халықаралық ғылыми конференциясының
БАЯНДАМАЛАР ЖИНАҒЫ**

**СБОРНИК МАТЕРИАЛОВ
XI Международной научной конференции
студентов и молодых ученых
«Наука и образование - 2016»**

**PROCEEDINGS
of the XI International Scientific Conference
for students and young scholars
«Science and education - 2016»**

2016 жыл 14 сәуір

Астана

ӘӨЖ 001:37(063)

КБЖ 72:74

F 96

F96 «Ғылым және білім – 2016» атты студенттер мен жас ғалымдардың XI Халық. ғыл. конф. = XI Межд. науч. конф. студентов и молодых ученых «Наука и образование - 2016» = The XI International Scientific Conference for students and young scholars «Science and education - 2016» . – Астана: <http://www.enu.kz/ru/nauka/nauka-i-obrazovanie/>, 2016. – б. (қазақша, орысша, ағылшынша).

ISBN 978-9965-31-764-4

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.

ӘӨЖ 001:37(063)

КБЖ 72:74

ISBN 978-9965-31-764-4

©Л.Н. Гумилев атындағы Еуразия
ұлттық университеті, 2016

Список использованных источников

1. Тропченко А.Ю. Методы вторичной обработки и распознавания изображений / А.Ю. Тропченко А.Ю., А.А. Тропченко - Изд-во: Университет ИТМО, 2015. – 215 с.
2. Viola P. Rapid Object Detection using a Boosted Cascade of Simple Features / P. Viola, M.J. Jones // proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2001), 2001
3. Viola P. «Robust real-time face detection / P. Viola, M.J. Jones // International Journal of Computer Vision, vol. 57, no. 2, 2004., pp.137–154
4. Гонсалес Р. Цифровая обработка изображений / Р.Гонсалес, Р.Вудс, ISBN 5-94836-028-8, изд-во: Техносфера, Москва, 2005. – 1072 с.

УДК 004.82; 004.912; 81.322.2

ОНТОЛОГИЧЕСКАЯ МОДЕЛЬ КАЗАХСКО-ТУРЕЦКОЙ И ТУРЕЦКО-КАЗАХСКОЙ СИСТЕМЫ МАШИННОГО ПЕРЕВОДА

Узгенбаева Райхан¹, Жеткенбай Лена²

¹Студент 4го курса «5B060200-Информатика», ЕНУ им. Гумилева, Астана, Казахстан

²Докторант 2го курса «6D060200-Информатика», ЕНУ им. Гумилева, Астана, Казахстан

Научный руководитель – Бекманова Г.Т., к.т.н., доцент

В условиях бурного развития технологий, когда информатизация касается каждой области повседневной жизни человека, с каждым днем увеличивается объем мировой коммуникации. Межкультурное общение становится неотъемлемой частью жизни многих людей, в этих условиях использование машинного перевода в сети Интернет – это повседневная реальность практически каждого активного пользователя сети. В настоящее время во всем Мире бурно развиваются технологии машинного перевода языков развитых стран Мира. Например, существуют эффективные технологии машинного перевода такие как GoogleTranslate, в меню которого доступны 80 языков; подобные сервисы, но с меньшими возможностями представляют Яндекс.Перевод, Translate.ru, BingTranslator, Apertium, BabelFish, Dicto, PROMT, GoldenDict, StarDict, WikiBhasha, SYSTRANet, FreeTranslation, BabelFish, Worldlingo, InterTran, ImTransator, Translate Online.ua, WindowsLifeTranslator [1]. В условиях, когда мировые лидеры в области машинного перевода не предоставляют сервисов качественного перевода с/на казахский язык, назрела необходимость в создании интеллектуального семантического переводчика тюркских языков на примере казахского и турецкого языка.

Впервые идея применения специальных машин для перевода слов с одного языка на другой была предложена в XVII в. знаменитыми математиками, естествоиспытателями и философами Готфридом Вильгельмом Лейбницем и Рене Декартом. Считается, что как самостоятельное научное направление научный перевод зародился, в марте 1947 г., когда криптограф Уоррен Уивер, руководивший отделением естественных наук Рокфеллеровского фонда, обратился с письмом к Норберту Винеру, в котором высказал мысль о рассмотрении задачи перевода с одного языка на другой как новую область использования технологий декодирования. [2]. За последние 50 лет машинный перевод бурно развивался и совершенствовался, сегодня большинство машинных переводчиков доступны онлайн.

Казахский язык относится к группе тюркских языков и имеет много общего с турецким языком. Это позволяет создавать машинные технологии семантического перевода с одного близкородственного языка на другой. Подобные переводчики являются более эффективными. Взаимоотношения между Республикой Казахстаном и Турцией во всех сферах традиционно продолжают развиваться, то есть необходимость в создании эффективного переводчика между

казахским и турецким языками и наоборот. В настоящее время в Республике Казахстан развиваются технологии машинного перевода, такие казахско-русские, русско-казахские, казахско-английские и т.д. переводчики [4-5], но казахско-турецкого и турецко-казахского машинного перевода в настоящее время не существует.

Для тюркских языков достаточно хорошо поддающихся формализации, поскольку с точки зрения морфологии в них строго определен порядок присоединения аффиксов, а с точки зрения синтаксиса строго определен порядок слов в предложении, достаточно хорошие результаты могут давать системы на основе грамматических правил. Статистические системы дают хорошие результаты там, где существуют большие текстовые корпуса. Но наибольшую эффективность машинного перевода дают гибридные системы с учетом семантики текста.

Нами выбрана предметная область «Туризм», поскольку туристические отношения между Казахстаном и Турцией развиты, а данная область понятна и может быть использована в качестве примера.

В качестве инструмента моделирования предметной области выбран редактор онтологий Protégé.

Protégé – это свободный, открытый редактор онтологий и фреймворк для построения баз знаний.

Платформа Protégé поддерживает два основных способа моделирования онтологий посредством редакторов Protégé-Frames и Protégé-OWL. Онтологии, построенные в Protégé, могут быть экспортированы во множество форматов, включая [RDF \(RDF Schema\)](#), [OWL](#) и XML Schema.

Protégé имеет открытую, легко расширяемую архитектуру за счёт поддержки модулей расширения функциональности [3].

Онтология является мощным и широко используемым инструментом для моделирования отношений между объектами, принадлежащими к различным предметным областям. Можно классифицировать онтологии на основе степени зависимости от задачи или прикладной области, модель представления онтологических знаний и выразительности, а также других критериев. Основная часть формально представленных знаний основана на концептуализации: объекты, концепции и другие объекты, которые считаются существующими в некоторой области интересов и отношения, которые держат среди них (Генесерет&Nilsson, 1987). Концептуализация является абстрактным, упрощенным взглядом на мир, который мы хотим представить для какой-то цели. Каждая база знаний – это система, основанная на знаниях, стремящаяся к некоторой концептуализации, явно или неявно. Онтология явная спецификация концептуализации. Термин заимствован из философии, где онтология является систематическим изложением Бытия. Для систем искусственного интеллекта, тем, что "существует" является то, что можно представить.

Когда знания предметной области представлены в декларативном формализме, множество объектов, которые могут быть представлены называется универсум дискурса. Этот набор объектов, и описываемые отношения между ними, отражаются в изобразительной лексике, с помощью которого программа, основанная на знаниях представляет знания. Таким образом, в контексте искусственного интеллекта, мы можем описать онтологию программы путем определения набора репрезентативных терминов. В такой онтологии определения связывают имена сущностей во вселенной дискурса (например, классы, отношения, функции или другие объекты) с текстом читабельной, описывающие то, что означают имена и формальные аксиомы, которые ограничивают интерпретацию и хорошо сформированные использование этих терминов. Формально онтология является утверждением логической теории [4, 5].

Прикладные онтологии описывают понятия, которые зависят как от задачи и предметной области онтологии. Применяемые онтологии основывается на общих принципах построения онтологии и семантических гипер-графов, используемых в качестве моделей для представления

знаний. Этот формализм определяет онтологию O как тройка (V, R, K) , где V представляет собой набор понятий данной предметной области, R представляет собой совокупность отношений между этими понятиями, и K представляет собой набор имен понятий и отношений в домен [6,7,8].

Онтологическая модель предметной области «Туризм» на турецком языке представлено на рисунке 1:

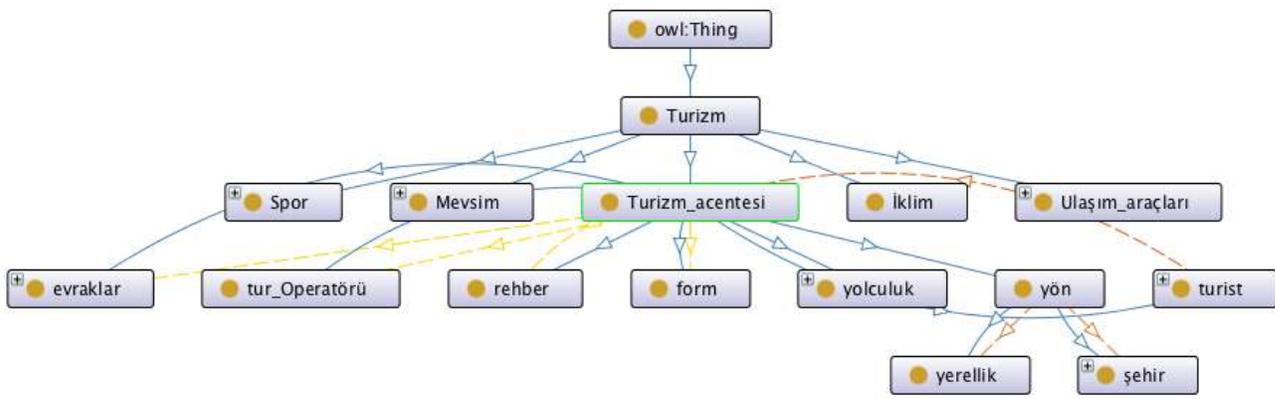


Рисунок 1 – Онтологическая модель предметной области «Туризм» на турецком языке.

Идентичная онтология на казахском языке представлено на рисунке 1:



Рисунок 2 – Онтологическая модель предметной области «Туризм» на казахском языке.

Более подробно на рисунках 3 и 4 показана онтологическая модель объекта **город**.

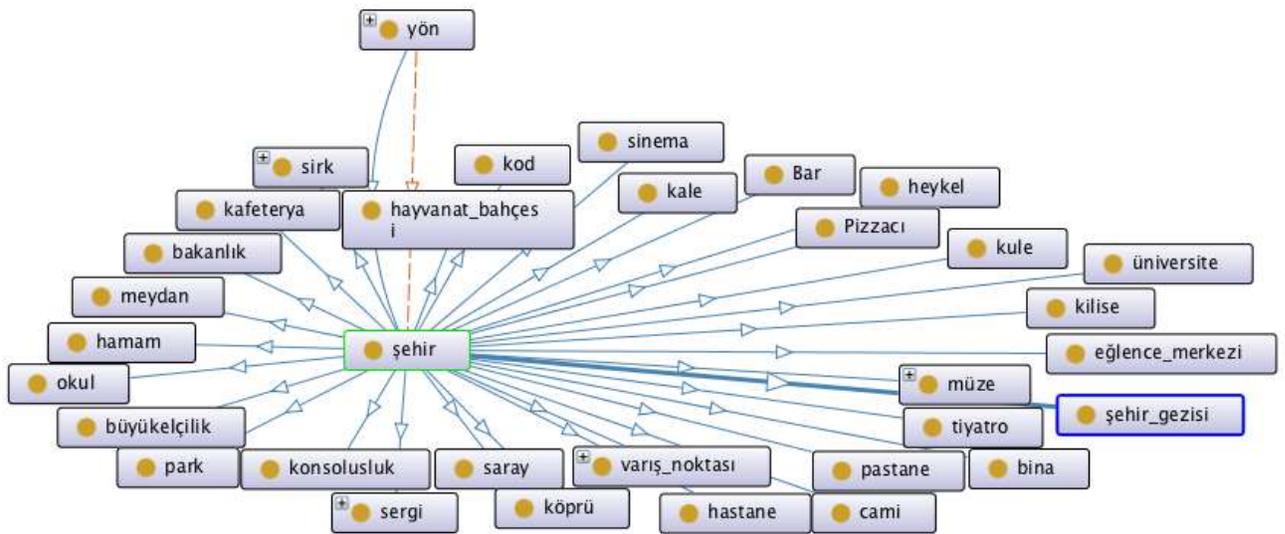


Рисунок 3 –Онтологическая модель «şehir»

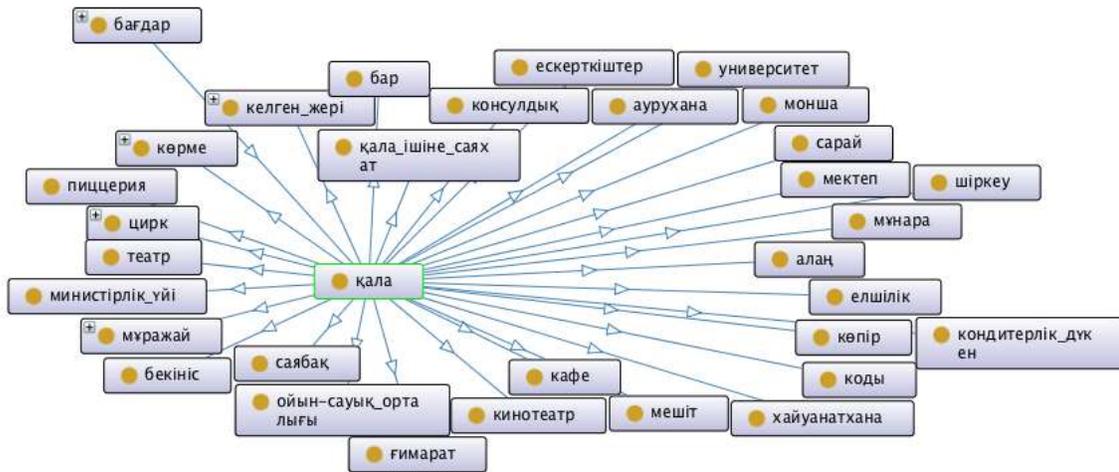


Рисунок 4 – Онтологическая модель «қала»

Ниже показана онтологическая модель объекта «водный транспорт»:

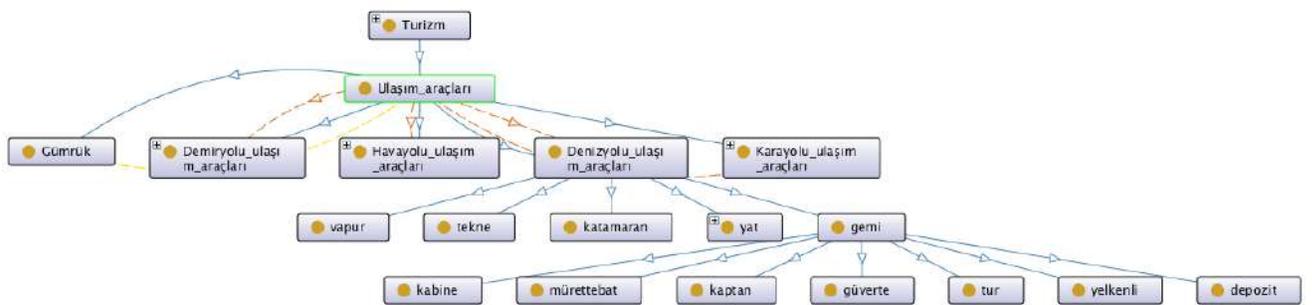


Рисунок 5 – Онтологическая модель «водный транспорт» на турецком языке

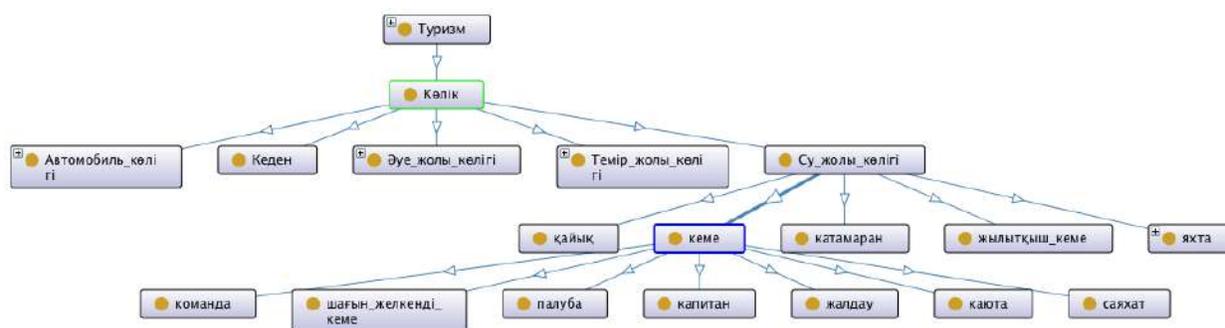


Рисунок 6 – Онтологическая модель «водный транспорт» на казахском языке

Полученная онтологическая модель предметной области позволяет формализовать ее знания и представить онтологии как ее спецификацию. Поскольку онтология предметной области Туризм на казахском языке симметрична предметной области Туризм на турецком языке это позволит создать универсальную базу знаний для семантического машинного перевода, вопросно-ответных систем и других систем, в которых в основе лежит семантическая обработка текста на естественном языке. А то обстоятельство, что объектами формализации являются родственные тюркские языки открывает потрясающие возможности обеспечения высокого качества машинного перевода.

Список использованной литературы

1. <https://811mtk.wordpress.com/2012/10/10>
2. <https://sites.google.com/site/sashakhts/istoria-masinnogo-perevoda>
3. <https://ru.wikipedia.org/wiki/Prot%C3%A9g%C3%A9>
4. Gruber T.R. A Translation Approach to Portable Ontology Specifications / Gruber T.R.// Knowledge Acquisition,1993, P.199-220
5. Gruber T.R. Toward Principles for the Design of Ontologies Used for Knowledge Sharing / Gruber T.R. // International Journal Human-Computer Studies. – 1995, - Vol. 43 - P.907-928
6. BanuYergesh, AsselMukanova, AltynbekSharipbay, GulmiraBekmanova, and BibigulRazakhova. Semantic Hypergraph Based Representation of Nouns in the Kazakh Language. Computación y Sistemas Vol. 18, No. 3, 2014 pp. 627– 635 ISSN 1405-5546 DOI: 10.13053/CyS-18-3-2041.
7. Mukanova, A., Yergesh, B., Bekmanova, G., Razakhova, B., Sharipbay, A. Formal models of nouns in the Kazakh language. Leonardo Electronic Journal of Practices and Technologies.
8. L.Zhetkenbay, A.A.Sharipbay, G.T.Bekmanova, M.Khabylashimuly, U.Kamanur. The semantical, ontological models and formalization rules Kazakh compound words. Turklang'14 II International Conference on Computer processing of Turkic Languages, 2014 .–Istanbul, 2014. – P.107-113. ISBN No.:978-975-561-449-6