

Семантический компонент в системах машинного перевода: обзор литературы.

Тюрина Людмила Сергеевна

Студент (магистр)

Московский государственный университет имени М.В.Ломоносова, Факультет иностранных языков и регионоведения, Москва, Россия

E-mail: tyurina.lyudmila.sergeevna@yandex.ru

Данная работа посвящена проблеме сохранения семантической информации в процессе машинного перевода. Статья представляет собой обзор актуальных на сегодняшний день работ, рассматривающих способы передачи семантического компонента в системах МП.

Основная задача работы - синтезировать данные по этому вопросу, чтобы получить представление о текущем положении дел в данной области и определить, какое из направлений является наиболее многообещающим для англо-русского перевода с целью дальнейших исследований в этой области.

На сегодняшний день уже достигнуты значительные успехи в развитии систем МП. Благодаря использованию грамматического и синтаксического анализаторов, активная разработка которых велась в 70е - 80е годы, удалось добиться значительных успехов в деле автоматического перевода текстов. Сегодня системы МП достаточно развиты для того, чтобы использоваться в промышленных масштабах, однако для улучшения качества перевода необходимо задействовать семантический компонент. Как научить компьютер "понимать" текст? На этим вопросом трудятся специалисты в области компьютерной обработки текстов на естественных языках в разных странах мира.

Данная работа логически разделена на 5 частей.

Для начала, необходимо очертить проблему, дав точное определение того, что такое "семантический компонент". Для этого мы обратимся к собственно лингвистическим теориям, рассмотрим термин "семантика", а также определения таких ключевых понятий как "значение" и "содержание" в рамках компьютерной лингвистики.

Далее рассмотрим основные проблемы, возникающие при попытке извлечения семантической информации из текста и ее перевода на другой язык. Сюда относятся проблемы связи семантики и синтаксиса, вопрос о необходимости создания универсальной схемы перекодирования, или же целесообразности адаптации каждой системы к конкретной паре языков и т.д. Данная часть в основе своей опирается на работы Н.Н. Леонтьевой.

В следующих двух частях рассмотрены современные подходы к передаче семантического компонента в системах МП, такие как метод семантических сетей, теория "Смысл-Текст", использование системы Abstract Meaning Representation Bank, а также упрощенная система семантического парсинга. В основе лежит теория И. А. Мельчука, а также работы европейских и американских специалистов в области компьютерной лингвистики - D. Santos, J. Andreas, B. Jones, D. Bauer, S. Clark. Ph. Koehn.

Завершающая часть работы призвана резюмировать проанализированные данные и сделать прогноз того, какой из предложенных методов автоматической обработки семантической информации может стать наиболее продуктивным для перевода между английским и русским языками. В качестве обоснования прогноза планируется анализ результатов работы уже существующих систем МП в соответствии с рассмотренными выше теоретическими концепциями.

Источники и литература

- 1) Автоматическая обработка текстов на естественном языке и компьютерная лингвистика : учеб. пособие / Большакова Е.И., Клышинский Э.С., Ландэ Д.В., Носков А.А., Пескова О.В., Ягунова Е.В. – М.: МИЭМ, 2011. – 272 с.
- 2) Леонтьева Н.Н. Автоматическое понимание текстов: Системы, модели, ресурсы. Академия, 2006. – 304 с.
- 3) Семенов Ю. А. Семантические сети [URL: http://book.itep.ru/10/sem_net.htm]
- 4) Santos, Diana. Semantics and (machine) translation, 1998. [URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.44.88&rep=rep1&type=pdf>]
- 5) Semantics-based machine translation with hyperedge replacement grammars. Bevan Jones*, Jacob Andreas*, Daniel Bauer*, Karl Moritz Hermann*, Kevin Knight. In Proceedings of COLING 2012. pp. 1359–1376. The COLING 2012 Organizing Committee. Mumbai, India. (*Joint first author in randomized order) [URL: <http://aclweb.org/anthology/C/C12/C12-1083.pdf>]
- 6) Semantic parsing as machine translation Jacob Andreas, Andreas Vlachos and Stephen Clark. ACL 2013. [URL: http://www.cs.berkeley.edu//papers/avc_smt_semparse.pdf]