**ОТЗЫВ**

о выпускной квалификационной работе

Попова Андрея Михайловича

«**Проблемы автоматического выравнивания**

**в параллельных корпусах флективных языков**»

Выпускная квалификационная работа А. М. Попова посвящена актуальной проблеме — выравниванию параллельных текстов на двух уровнях — на уровне предложений и на уровне документов — на примере трех языков. Полученные автором результаты могут найти применение в разных приложениях, связанных с автоматической обработкой текста (например, при создании систем машинного перевода), а также будут востребованы переводчиками и лексикографами.

Работа четко структурирована и состоит из введения, трех глав, заключения и трех приложений.

В первой главе обсуждаются теоретические основы работы – корпусные технологии, параллельные корпусы текстов, вопросы их разметки и способы представления.

Во второй главе обосновывается выбор материала для исследования (англо-русские, русско-чешские и чешско-английские языковые пары), детально описываются методы выравнивания параллельных текстов, в том числе предлагаемый автором алгоритм для выравнивания по документам. Данный алгоритм построен на основе подсчета ключевых слов (под ними понимаются низкочастотные слова), лемматизации и двуязычного словаря.

Практическая часть работы, представленная в третьей главе, посвящена тестированию разработанной в рамках исследования программы и экспериментам с привлечением существующих инструментов. Автором была произведена оценка выравнивания по предложениям при помощи программы LF Aligner и рассмотрено использование лемматизации для улучшения качества результатов. Также автором оценивается собственный алгоритм выравнивания текстов по документам, для которого был разработан программный инструмент на языке C# (его код приведен в приложениях 1 и 2).

Хотелось бы высказать ряд соображений, возникших при знакомстве с работой.

1. Возможно, следовало бы более четко сформировать новизну исследования.
2. Желательно было бы привести более строгое определение для термина «выравнивание» (стр. 8).
3. Представляется слишком категоричным утверждение о том, что топонимы и имена собственные «нередко пишутся одинаково или очень схожим образом в чешском и английском» (стр. 41). Например, Rakousko и Austria (Австрия), Benátky и Venice (Венеция), Řeckoи Greece (Греция) и др.
4. На чем был основан выбор программы LF Aligner для проведения экспериментов?

Отмеченные моменты не влияют на качество работы в целом и носят дискуссионный характер.

Работа выполнена на высоком уровне, что характеризует автора как высококлассного специалиста, способного к самостоятельной научной деятельности. Выпускная квалификационная работа А. М. Попова представляет собой законченное исследование и отвечает всем требованиям, предъявляемым к работам подобного рода, и заслуживает самой высокой оценки.



Старший преподаватель

кафедры математической лингвистики

филологического факультета СПбГУ,

к.ф.н. Хохлова М. В.

02.06.2014