

### 0.1. *Бакиева А.М. Подходы к созданию моделей определения тем текстов на тюркских языках*

Задача определения тем текстов напрямую связана с задачей автореферирования. Последняя довольно успешно может решаться при помощи тезаурусов. Можно выделить следующие новые задачи, связанные с компьютерным реферированием

1. Создание одноязычных рефератов из источников на разных языках из одного источника при наличии тезауруса;
2. Построение рефератов по гибридным источникам, включающим как текстовые, так и числовые данные в разных формах (таблицы, диаграммы, графики и т.д.);
3. Создание рефератов на основе массивов документов. Например, построение единого реферата по сборнику тезисов докладов научной конференции [2].

При создании системы определения тем текстов важным этапом является семантический анализ. Даже частичная его реализация могла бы повысить эффективность работы такой системы. В качестве инструмента, позволяющего производить синтактико-семантический анализ предложений нами был выбран Link Grammar Parser [1].

В парсере используется около 100 основных связей, некоторые из них разбиваются на 3-4 подслучая, среди них, например, выделяются такие семантические группы, как ситуационные наречия, наречия времени, вводные наречия и др. В процессе разбора рассматриваются все варианты связей между словами, и выбираются среди них те, которые удовлетворяют критерию проективности (связи не должны пересекаться) и критерию минимальной связности (получившийся граф должен содержать наименьшее число компонент связности). Парсер позволяет анализировать большое число конструкций, включая многочисленные редкие выражения и идиомы [3].

Точность работы Link Grammar во многом зависит от полноты подключаемых словарей. На сегодняшний день мы уже исследовали возможность подключения и создания словарей на казахском и турецком языках. Поскольку эти языки являются агглютинативными, приходится учитывать их морфологические особенности.

В настоящее время мы выделили оптимальный набор связей, подходящий для работы с тюркскими языками. Например: АІ - Соединяет существительное (в роли подлежащего / дополнения) с прилагательным (в роли определения), стоящим перед ним; ЕІ - Соединяет глагол с предыдущим наречием (в роли обстоятельства); ІІ - Соединяет существительное (в роли дополнения) с последующим предлогом и т.д. Но еще имеются некоторые недоработки по

окончаниям и суффиксам слов.

## Список литературы

- [1] БАТУРА Т.В., МУРЗИН Ф.А. Машинно-ориентировочные логические методы отображения семантики текста на естественном языке / Новосибирск: Наука, 2008. — 114 с.
- [2] СОЛТОН ДЖ. Динамические библиотечно-информационные системы // Издательство «Мир». — 1979. — С. 41–50.
- [3] БАТУРА Т.В., МУРЗИН Ф.А., ПЕРФИЛЬЕВ А.А., ШМАНИНА Т.В. Методы повышения эффективности поиска информации на основе синтаксического анализа: Монография / Новосибирск: Изд-во СО РАН, 2014. — 45 с.
- [4] ХАН У., МАНИ И. Системы автоматического реферирования // Открытые системы, 2000. — 67 с.