



ПРОЦЕДУРА ФОРМИРОВАНИЯ ИНВАРИАНТНОЙ РЕПРЕЗЕНТАЦИИ СИТУАЦИИ ДЛЯ АВТОМАТИЗАЦИИ ПРОЦЕССА ПОНИМАНИЯ ТЕКСТА В СИСТЕМЕ АВТОМАТИЧЕСКОГО РЕФЕРИРОВАНИЯ

Лазаренко О.В.

*Харьковский гуманитарный университет «Народная украинская академия»
г. Харьков, ул. Лермонтовская, 27, тел. 050-300-82-83
e-mail: lazovlad@yandex.ru*

Работа над созданием искусственного интеллекта (ИИ) длится не первое десятилетие. Однако ключевая задача - научить компьютеры работать подобно человеческому мозгу - так же далека от своего решения, как и раньше. Одна из причин такого положения состоит в том, что разработчики ИИ, по мнению известного разработчика компьютеров в Силиконовой долине Джеффа Хокинса, «хотят достичь поставленной цели, обойдя вниманием вопрос о сути разума, о том, что означает слово “понимать”... Этим самым они “выплеснули с водой ребенка” — создавая мыслящие механизмы, забыли о разуме! Но все попытки создания искусственного интеллекта без учета особенностей естественного обречены на провал» [1].

Сравнивая работу человеческого мозга с работой компьютера, Дж. Хокинс задался вопросом - *какая составляющая разума отсутствует в компьютере?*

В поисках ответа на этот вопрос в августе 2002 года он открыл научно-исследовательский центр по изучению мозга, в котором первостепенное значение уделили изучению неокортекса — части головного мозга человека, ответственной за интеллект [1].

В ходе изучения неокортекса были обнаружены определенные особенности его работы, среди которых для нас представляет особый интерес способность мозга использовать инвариантные репрезентации объекта с сохранением его наиболее важных признаков на основе относительных измерений, пропорций и других характеристик, в которых возможны существенные упущения в сравнении с конкретным образом. Оказалось, что мозг запоминает важные взаимосвязи внешнего мира, а не привязывается к отдельным его элементам.

Аналогичным образом действует человеческий мозг и при чтении текстовой информации. Согласно гипотезе, выдвинутой голландским ученым А. ван Дейком о том, что при чтении текста люди часто обрабатывают информацию не полностью или неточно и, тем не менее, понимают текст. «Языковому пользователю нет необходимости дожидаться конца абзаца, главы или целого текста, чтобы понять, о чем идет речь в тексте или в его фрагменте, ... пользователь языка может догадаться о теме текста уже после минимума текстовой информации из первых пропозиций. Догадку может подтвердить самая различная информация: заглавие, тематические слова, тематические первые предложения...» [2], то есть наиболее важные аспекты и их взаимосвязи.

В своих исследованиях процесса понимания текста [3] мы пришли к необходимости разработки ситуационных моделей, позволяющих уйти от использования онтологий и упростить автоматизацию процесса понимания текстов.

В разрабатываемой нами системе автоматического реферирования ситуационная модель формируется в виде накопителя текстовых баз определенной тематики, автоматически извлекаемых из текста в процессе его смыслового анализа в соответствии с разработанным алгоритмом извлечения основных смысловых аспектов текста.

Полученная таким образом ситуационная модель является основой для создания *инвариантной репрезентации ситуации*, представляющей собой набор наиболее важных признаков, выделенных на основе относительных характеристик ситуации, в которых возможны существенные упущения в сравнении с конкретной ситуацией, описываемой в конкретном тексте.

Возникает вопрос: какие признаки считать наиболее важными для описания данной ситуации, чтобы включить их в инвариантную репрезентацию?

Существенную помощь в этом могут оказать заголовки статей как концептуальные инварианты текстов, на базе которых создавалась данная ситуационная модель.

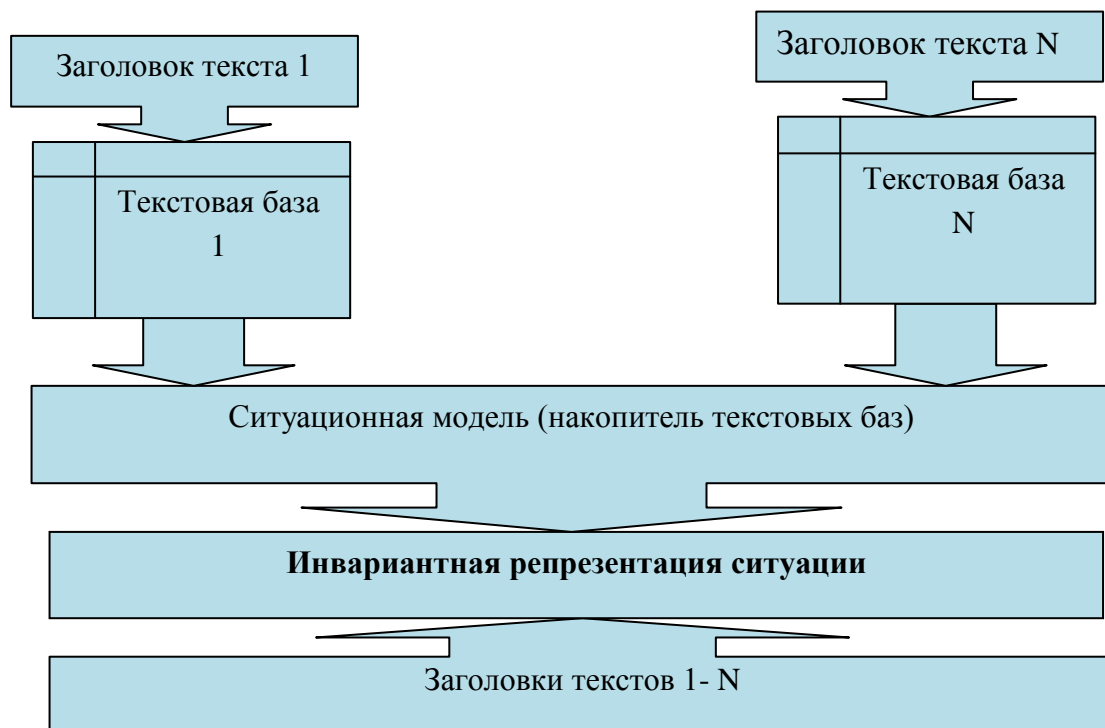


Рис. 1. Процедура разработки инвариантной репрезентации ситуации в системе автоматического реферирования

Таким образом, с использованием текстовых баз, задающих контекстную семантику, и формируемых на их основе ситуационных моделей, содержащих информацию, актуализируемую в процессе понимания текста, а также заголовков всех текстов можно выделить наиболее важные смысловые составляющие определенной ситуации, которые и составят инвариантную репрезентацию ситуации.



Может показаться, что все это уводит нас от основной задачи исследования. Однако это не так. Чем глубже и яснее мы представляем процесс понимания, осуществляемый человеком, тем точнее будет настройка системы автоматического реферирования на анализ смысла исходного текста. Помимо этого исследование и моделирование некоторых аспектов процесса понимания открывают нам тайны работы мозга. Мы видим, как работы в разных областях, таких как исследование механизмов работы мозга (Дж. Хокинс), разработка стратегий понимания дискурса (А. ван Дейк) и моделирование процесса реферирования (Лазаренко О.В.) сошлись в одной точке – инвариантных представлениях, лежащих в основе указанных процессов. Результаты этих исследований подтверждают прямо и косвенно тот факт, что мозг в процессе распознавания объекта, фактов, ситуаций и проч. вспоминает важные взаимосвязи внешнего мира, а не привязывается к отдельным его элементам.

Все вышеизложенное позволяет надеяться, что разрабатываемая нами процедура смыслового анализа текста позволит обеспечить более качественный результат автоматического реферирования за счет использования доступа к информации необходимой для синтеза реферата путем:

1. выделения макроструктуры текста и формирования на ее основе текстовой базы;
2. формирования в автоматическом режиме ситуационных моделей в виде накопителей текстовых баз, используемых для более точного понимания смысла конкретного текста;
3. извлечения знания, имеющегося в тексте, через актуализацию его с помощью инвариантных репрезентаций ситуаций.

Список литературы

1. *Хокинс Дж., Блейкли С.* Об интеллекте / Дж., Хокинс, Блейкли С. - М.: Издательский дом "Вильямс", 2007. - 240 с.
2. *Дейк ван Т. А.* Стратегии понимания связного текста / Т. А. ван Дейк, В. Кинч // Новое в зарубежной лингвистике. – Вып. 23: Когнитивные аспекты языка. – М., 1988. – С. 153–211.
3. *Лазаренко О.В.* Разработка интеллектуальной системы автоматического реферирования с использованием текстовых баз и ситуационных моделей /О.В. Лазаренко // MegaLing'2013. Горизонти прикладної лінгвістики та лінгвістичних технологій : доп. міжнар. наук. конф., Україна, Київ, 20-23 листопаду 2013 г.