

можно назвать: устранение повторов, вставных конструкций, более короткий перифраз, использование аббревиатур и композитов, синтаксический эллипсис. Ко второму виду компрессии относится одновременное сокращение языкового кода и объема семантической информации с возможностью полного восстановления семантического наполнения текста. Третий вид компрессии – это одновременное сокращение языкового кода и объема семантической информации (компрессия информации) без возможности восстановления информации полностью без обращения к первичному тексту.

Список литературы

1. Дридзе Т.М. Текстовая деятельность в структуре социальной коммуникации. Проблемы семисоциопсихологии: монография. – М.: Наука, 1984. – 232с.
2. Уланович О.И. Текст и его понимание // Вестник МГЛУ. Серия 2. Психология, педагогика, методика преподавания иностранных языков. – № 3. – Минск: МГЛУ, 2001. – С. 22 – 30.
3. Жеребило Т. В. Словарь лингвистических терминов. – Изд. 5-е, испр. и доп. – Назрань: ООО «Пилигрим», 2010. – 486 с.
4. Валгина Н.С. Теория текста: учебное пособие. – Москва: изд-во МГУП «Мир книги», 1998. – 210 с.
5. Бельский А. Извлечение информации из неструктурированных текстов [Текст] / А. Бельский // КомпьютерПресс, 2008. – №2. – С. 74-79.

ТЕХНОЛОГИИ ОБРАБОТКИ ИНФОРМАЦИИ И МЕТОДЫ АВТОМАТИЗИРОВАННОГО РЕФЕРИРОВАНИЯ И АННОТИРОВАНИЯ

Жилинская О.И., Чопорова Е.И.

Воронежский институт МВД России, Воронеж,
e-mail: choporova_ekaterina@mail.ru

Появление новейших информационных технологий предложило обществу безграничные возможности в сфере извлечения, обработки и применения информации. Вследствие этого информация превращается в один из важнейших ресурсов государства, а аннотация и реферат становятся важнейшими способами, которые способны обеспечивать эффективный обмен извлеченной информацией, сокращая, таким образом, время, отводимое на знакомство с новыми данными и их обработку.

Сущность реферирования и аннотирования, как основных лингвистических средств вербального сжатия информации с сохранением ее основного значения, заключается в максимальном сокращении объема информации при существенном сохранении ее основного содержания.

Аннотация (от лат. *annotatio* – замечание) представляет собой краткую характеристику смыслового компонента печатного произведения, и, как правило, содержит предельно краткую характеристику первоисточника описательного типа. В аннотации в обобщенном виде раскрывается тематика публикации без полного отражения ее содержания и дается ответ на вопрос, о чем именно говорится в первичном источнике информации. Композиционно аннотация представляет собой текстовый документ, состоящий из двух частей, где в первой части сформулирована основная тема первоисточника, а во второй – основные положения аннотируемого текста [1].

Реферат (от лат. *refero* – сообщая) является кратким письменным изложением содержания научного труда (трудов) по определенной теме с раскрытием его (их) основного содержания по всем затронутым вопросам, и, сопровождаемое оценкой и выводами референта. Как правило, реферат должен дать читателю объективное представление о характере освещаемой работы и сообщить наиболее существенные моменты ее содержания [2].

В отличие от аннотации в реферате не только освещается основная тема первичного документа, но и достаточно подробно раскрываются основные поло-

жения реферируемого первоисточника. Также реферат предоставляет описание первичного документа, оповещает о выходе в свет и о наличии соответствующих первичных документов, а также является источником для получения справочных данных и самостоятельным средством научной информации [3].

Составление аннотаций и рефератов человеком (референтом) требует от него значительных умственных усилий, специальных знаний и умений, связанных со сжатием текста, а также специальных лингвистических навыков перефразирования и т.д., в связи с чем, необходимость создания эффективных методов автоматизированного реферирования и аннотирования уже долгое время является достаточно актуальной.

Процесс автоматизированного сжатия информации на естественном языке осуществляется на основе компьютеризированных систем, и соответственно, влечет за собой создание специальных инструментов в двух основных областях знаний – прикладной лингвистике и системном анализе и обработке данных. Прикладная лингвистика занимается автоматизацией интеллектуальной деятельности, связанной с использованием языка, а именно: созданием систем искусственного интеллекта (автоматизированных) информационно-поисковых систем, систем автоматического аннотирования и реферирования информации [4].

В области системного анализа автоматизированная обработка текста осуществляется в рамках текстомайнинга (text mining), который также известен как текстовый дейтамайнинг (text data mining). Текстомайнинг является более узким понятием, так как предназначен для извлечения новых знаний из неструктурированных текстовых массивов, в то время как дейтамайнинг позволяет находить новую информацию в больших объемах структурированной информации, т.е. хранимой в базах данных. Таким образом, после того, как неструктурированные текстовые массивы переводятся в структурированные, данные могут обрабатываться с помощью стандартных методов дейтамайнинга [5].

Одной из первых работ в области автоматизированного сжатия текста на естественном языке посредством компьютеризированных систем считается статья Ханса Петера Луна (H. P. Luhn) «The automatic creation of literature abstracts», опубликованная в 1958 году в научном журнале «IBM Journal of Research and Development» [6]. Начиная с данной точки отсчета предлагалось множество подходов к автоматизированному аннотированию и реферированию, некоторые из которых основывались на выявлении определенных статистических закономерностей распределения терминов в тексте или их взаимного расположения в нем [7; 8], в то время как другие использовали внутренние структуры текста, т.е. ориентировались на выявление ключевых предложений и ключевых «узлов» первоисточника [9].

В настоящее время выделяют два основных подхода к автоматизированной аннотированной компрессии текстовых документов: извлечение информации или экстракция и обобщение или абстракция [10; 11].

Извлечение предполагает выделение наиболее важных и существенных информационных блоков (фрагментов), чаще всего предложений из исходного текста и соединение их в реферат или аннотацию. Такое выделение ключевых предложений осуществлялось на основе позиционных и тематических критериев, т.е. на основе местоположения предложения в документе или на базе наличия ключевых слов в предложении. Подобные методы, основанные на этом подходе, зачастую называют поверхностными [10].

Обобщение предполагает использование предварительно разработанных грамматик естественных языков, тезаурусов, онтологических справочников и т.д., на основании которых выполняется переформулирование исходного текста и его обобщение. Обобщение при реферировании – это фактически процесс порождения или генерации нового текста, который характеризуется тремя основными этапами: адекватной перцепцией и анализом первоисточника с его дальнейшей девербализацией и созданием внутреннего представления его структуры (чаще всего, в виде семантических деревьев), дальнейшее семантическое сжатие внутреннего представления и синтез нового текста (реферата), или ревербализация экстрактированной внутренней структуры исходного текста. Методы, основанные на обобщении, считают более глубинными [10; 11].

Существует ряд специальных инструментов для работы с текстовой информацией, которые позволяют читателю достаточно быстро ознакомиться с текстами любой тематики. Среди них можно отметить такие программные продукты как *WordStart*, *Aero Text*, *STATISTICA Text Miner*, *Attensity suite*, «ГАЛАКТИКА».

Однако если говорить о применимости определенного программного продукта к текстам научно-технического характера, наиболее полезной программой представляется *TextAnalyst*, которая позволяет увидеть все понятия, употребляемые в предъявленных текстах, а также может выстроить все взаимосвязи и оценить их вес, представить содержание в виде иерархии тем и подтем.

TextAnalyst – это средство семантического анализа, навигации и поиска в неструктурированных текстах, результат совместного действия технологий лингвистического анализа и нейросетей. Данная система может быстро резюмировать документы в текстовой базе и классифицировать их в группы, а также облегчает поиск семантической информации либо может сфокусировать изучение текста на каком-то определенном предмете [5].

Одно из бесспорных преимуществ программы *TextAnalyst* – это возможность построения сети семантических связей текста (Semantic Network), которая в лингвистике текста известна также как денотатная карта, информационный граф, семантическое дерево текста и т.д. Полученная семантическая сеть служит основой для дальнейшего смыслового анализа текста. Семантическая сеть представляет собой наиболее важные или ключевые понятия, извлеченные из текста, с выражением взаимосвязей или отношений между ними, оцененных на основе их относительной значимости [5].

В основу данной программы заложен ряд основных принципов, в частности, принцип ассоциативности, который подразумевает такое хранение информации, когда при появлении небольшого ее фрагмента программа сразу отсылает читателя к тому месту, где она хранится. Это дает возможность статистической обработки данных: при совпадении фрагментов, система налагает их друг на друга, что позволяет установить частоту встречаемости различных элементов. В результате такого анализа формируется семантическая сеть – смысловая структура, характеризующая смысл текста, в которой понятия (слова и словосочетания) объединяются ассоциативными связями в соответствии с их совместной встречаемостью. При этом из семантической сети исключаются слова, являющиеся общеупотребительными и не несущие самостоятельной смысловой нагрузки.

Второй принцип учитывает особенности сна человека, когда информация, поступившая на хранение

в кору мозга, переупорядочивается в соответствии с ее связями. При этом весовые характеристики, сформированные на первом этапе анализа, преобразуются таким образом, что понятия, не часто встречавшиеся во входной информации, но связанные мощными связями с основными понятиями, в свою очередь, приобретают большой смысловой вес [12].

В дальнейшем сформированная семантическая сеть сама включается в работу по структуризации информации, относя фрагменты текста к соответствующим понятиям и связям сети. Таким образом создается так называемое тематическое дерево – представление структуры текста в виде многоуровневой иерархии тем и раскрывающих их подтем. В зависимости от смыслового строения текста, тематическое дерево может иметь один корень (главную тему), или расходиться на субдерева несвязанных друг с другом тем.

Семантическая сеть и тематическое дерево предъявляются для исследования пользователю *TextAnalyst*, который за каждым понятием и связью видит определенный смысл, заложенный в тексте. При этом пользователь избавляется от необходимости формирования модели текста, так как за него данное действие выполняется программой *TextAnalyst*.

Анализ текста, как правило, завершается автоматическим реферированием. При этом из текста отбираются фразы, содержащие наиболее весомые понятия с наиболее сильными связями, которые несут максимальную информацию о содержании текста. Фразы представляются в порядке их появления в исходном тексте [12].

Однако необходимо отметить, что реферат, получаемый на выходе при использовании автоматических программ обработки текстовой информации, выполнен по принципу экстракции или извлечения готовых фраз из текста-первоисточника, и, потому, не может в полной мере считаться рефератом или аннотацией искомого документа. Являясь незаменимым инструментом при составлении понятийных деревьев или ассоциативных карт предъявляемой информации, существующие на данный момент программы автоматизированной переработки текста, пока еще не могут в полной мере заменить человека при выполнении высокоорганизованной интеллектуальной деятельности. Сложнейшие операции, связанные с осмыслением, анализом, компрессией, языковым декодированием и перекодированием текста в настоящее время способен выполнить только человек, имеющий достаточно высокий уровень развития, интеллекта и языковой подготовки.

Тем не менее, современные системы реферирования способны оказать неоценимую помощь людям, чья профессиональная деятельность связана с анализом большого количества информации. У данного научно-инженерного направления совместно с различными областями прикладной лингвистики есть много перспективных путей развития.

Список литературы

1. Учебный словарь стилистических терминов [Электронный ресурс]. – (www.stilistics.academic.ru).
2. Методология истории: некоторые актуальные проблемы. Учебное пособие в помощь школьным учителям истории и обществознания, повышающим квалификацию в системе дополнительного профессионального образования / Е.Ю. Бобкова [и др.] – Самара: ПСГА. – 170 с.
3. Маркушевская Л.П., Цапаева Ю.А. Аннотирование и реферирование (Методические рекомендации для самостоятельной работы студентов) / Л.П. Маркушевская, Ю.А. Цапаева. – СПб ГУ ИТМО, 2008. – 51 с.
4. Энциклопедический словарь [Электронный ресурс]. – (www.dic.academic.ru).
5. Бельский А. Извлечение информации из неструктурированных текстов [Текст] / А. Бельский // КомпьютерПресс, 2008. – №2. – С. 74-79.

6. Luhn, H.P. (1958). The automatic creation of literature abstracts, IBM Journal of Research and Development, issue 2 / 1958, p. 159-165. – (<http://courses.ischool.berkeley.edu/i256/f06/papers/luhn58.pdf>).
7. Поспелов Д.А. Из истории искусственного интеллекта: история искусственного интеллекта до середины 80-х годов / Д.А. Поспелов // Новости искусственного интеллекта. – вып. 4. – 1994. – С. 70-90.
8. Гиляревский Р.С. Методы автоматизированного фрагментирования текста, отражающие на характеристике внутреннего состава фрагментов / Р.С. Гиляревский, С.И. Гиндин // Семиотика и информатика. М.: ВИНТИ, 1977. – Т.9. – С. 35-84.
9. Новиков А. И. Текст и его смысловые доминанты / Под ред. Н.В. Васильевой, Н.М. Нестеровой, Н.П. Пешковой. – М.: Институт языкознания РАН, 2007. – 224 с.
10. Технология анализа данных: Data Mining, Visual Mining, Text Mining, OLAP / А.А. Барсегян [и др.] – СПб.: БХВ-Петербург, 2007. – 384 с.
11. Браславский П., Колычев И. Автоматическое реферирование веб-документов с учетом запроса. Грант ООО «Яндекс» №102707, (compau@yandex.ru/grant/2005/11_Braslavski_102707.pdf).
12. [Электронный ресурс]. – (<http://www.analyst.ru/index.php?lang=rus&dir=content/tech/&id=approach>).

ПРЕИМУЩЕСТВА ОБУЧЕНИЯ ЗА РУБЕЖОМ И СЛОЖНОСТИ МЕЖДУНАРОДНОГО ОБМЕНА СТУДЕНТОВ

Щепетова А.М.

*Сибирский федеральный университет, Красноярск,
e-mail: anastasiashepetova@gmail.com*

Getting education abroad is very prestigious in Russia. Presence of a foreign university diploma or international appendix to a Russian university diploma is considered to be advantage for employment. It is believed that participation of Russia in the Bologna process where academic mobility is an important indicator should encourage international student exchange. Although training of Russian students at foreign universities is still more of an exception rather than a rule this trend is developing. In particular, at the beginning of 2014 Russia launched the program «Global Education». Its purpose is to create conditions for Russian students in the best universities of the world at the expense of the state budget. «Each student included in this program will receive 1.5 million rubles from the state annually». A very important requirement is after training specialists will have to return to work in Russia. If the specialist wishes to stay and work abroad he will be obliged not only to return the money spent on it but also to pay 200 % fine. According to UNESCO Statistics Institute the number of the students who go to study abroad from their country annually increases by 12%. Now the number of those being trained abroad reaches 3,5 million people. The most active are Chinese. According to the collected data of the Institute of International Education (IIE) over 440 thousand Chinese are currently being trained abroad. Among other countries students who get their education abroad most often are the citizens of India and South Korea. However their educational flows abroad are twice lower. As for Russians, the biggest part of international students from this country prefer to study in the USA. However, many of them go to get education in Germany, France or the UK. The advantages of studying abroad is promoted a lot by advertising. In these promotional materials the emphasis is put on the prestige and accessibility. Most universities that are mentioned in them are located in eastern Europe. Of course, the reliability of these sources of information can be doubted because of biased coverage of this issue. However, there is sufficient number of objective publications on the Internet and in press where one can find the analysis of the positive aspects of educational exchange with foreign universities. Studying abroad helps to develop independence, both financial and psychological because while your parents are away you can rely on yourself only. It encourages the development of self-sufficiency. It motivates to look for

learning options, applying different methods and tools for obtaining information. In most places class attendance is not controlled. So students can participate in one activity and the other one can be replaced by something more interesting. In my opinion «free» way of learning is not always an advantage. In particular, those students who are accustomed to more strictly regulated training schedule have difficulties in European universities with planning independent learning and choosing subjects and areas of training. When a student participates in exchange program the advantages of studying abroad include not only prestige and well-organized international procedure but also the absence of accessibility issue as the most costs are covered by sponsoring international organization. The main barrier preventing Russian students from learning efficiently is misunderstanding of foreigners, their language, culture, mentality. For example there can be different dialects in different parts of the country that make language understanding more difficult. Social adaptation of students in a foreign country is a very individual process. For some of them, this process takes just a few weeks, and for someone it lasts for several months or even a year. Another negative factor of Russian students adaptation is being homesick. It affects the learning process a lot («nostalgia is taking a person over so much that his thoughts are not about learning»). Another negative factor is that the average standard of living in most countries where Russian students participate in international exchange is higher than in Russia. Particularly students who came from the provincial universities feel this way. Expenses on food, clothing and entertainment abroad they usually are much larger than expected. It is also becoming a significant obstacle to adaptation. Among the other negative factors I can specify uncertainty of Russian students in their abilities, affordability of staying in another country, personal motivation to live there long enough and to be successful in training at a foreign university. A student who participates in the international exchange needs to get health insurance, apply for a visa, register in another country which requires a lot of effort and time. It seems quite difficult for many students. The Key Questions that students ask are: – Where will I live? – How will I communicate with foreigners? – What am I going to wear and eat? – Who will be the classmates and teachers? Fears and doubts, difficulties of Russian students adaptation abroad are to a large degree the result of the underdevelopment in the international students exchange practices in Russia. The lack of such experience in many Russian universities also causes problems for exchange students. The problem of entering the new sociocultural environment can be solved by strengthening language training and organizing some kind of studies of culture, customs and laws of different countries in Russian higher school when the universities have agreements to carry out exchange programs with these countries. In this case it is possible for the exchange to become mutually beneficial. It will help to establish strong links with the foreign partners. And the Russian university itself will attract more interest of foreign students. In this case it will be ready to accept them and provide them the same high-quality training and accommodation. Siberian Federal University is continuing to strengthen partnerships with foreign universities that allows it to organize various kinds of exchange programs as well as special training focusing on talented students of different majors. Organization of short courses about the the partners such as American University in Washington or Jena university or Polytechnic University of Madrid and their countries can make future exchange programs to be more beneficial for all the participants.