

О МОДИФИКАЦИИ СИСТЕМЫ ОТСЕЧЕНИЯ СЕМАНТИЧЕСКИ НЕРЕЛЕВАНТНОЙ И НЕПЕРТИНЕНТНОЙ ТЕКСТОВОЙ ИНФОРМАЦИИ

Булдакова А.А.* , Кудрявцев А.Г.

Уральский федеральный университет имени первого Президента России
Б.Н. Ельцина, г. Екатеринбург, Россия

*E-mail: ainiu@mail.ru

ABOUT THE MODIFICATION OF THE SYSTEM OF CUTTING SEMANTICALLY IRRELEVANT AND NON-PERTINENT TEXT INFORMATION OFF

Buldakova A.A., Kudryavtsev A.G.

Ural Federal University, Yekaterinburg, Russia

Annotation. There is the well-known idea how to cut semantically irrelevant and non-pertinent text information off. We carried out the structural modification description of this idea.

Известна система отсечения семантически нерелевантной и непертигентной текстовой информации [1]. Нами проведена ее модификация. Структурная модель предлагаемой системы – на рисунке 1.

На рисунке 1 были использованы следующие обозначения: 1 – преаратор (блок автоматического понимания текстов [2]); 2 – блок рубрицирования текстов по терминам [3]; 3 – блок взвешивания терминов; 4 – блок работы с семантическими связями; 5 – решатель; а – читаемый текст; б – библиотечные тексты; в – взвешенный пользовательский список рубрик; г – взвешенный пользовательский словник; д – взвешенный пользовательский список типовых семантических отношений [2]; е – пороговые значения оценок Q0 – Q3 (см. ниже); ж – неранжированный словник читаемого текста; з – таблица предложений читаемого текста; и – семантические структуры [2,4], обнаруженные в читаемом тексте; к – укороченный словник читаемого текста¹; л – вероятностные оценки встречаемости обнаруживаемых рубрикаций; м – оценка Q0 рубрикационного соответствия читаемого текста пользователю; н – взвешенный (по значимости терминов) словник; о – оценка Q1 терминологического соответствия читаемого текста пользователю; п – оценка Q2 соответствия читаемого текста пользователю по семантическим связям; р – оценка Q3 снятия семантической неопределенности [5,6] в читаемом

¹ Не исключено, что в процессе работы блока 2 не будут найдены рубрики по некоторым терминам читаемого текста, что приводит к формированию укороченного словника, не содержащего указанные термины.

тексте; с,т – стоп-сигналы для блоков, соответственно, 3 и 4; у – сообщения о несоответствии.

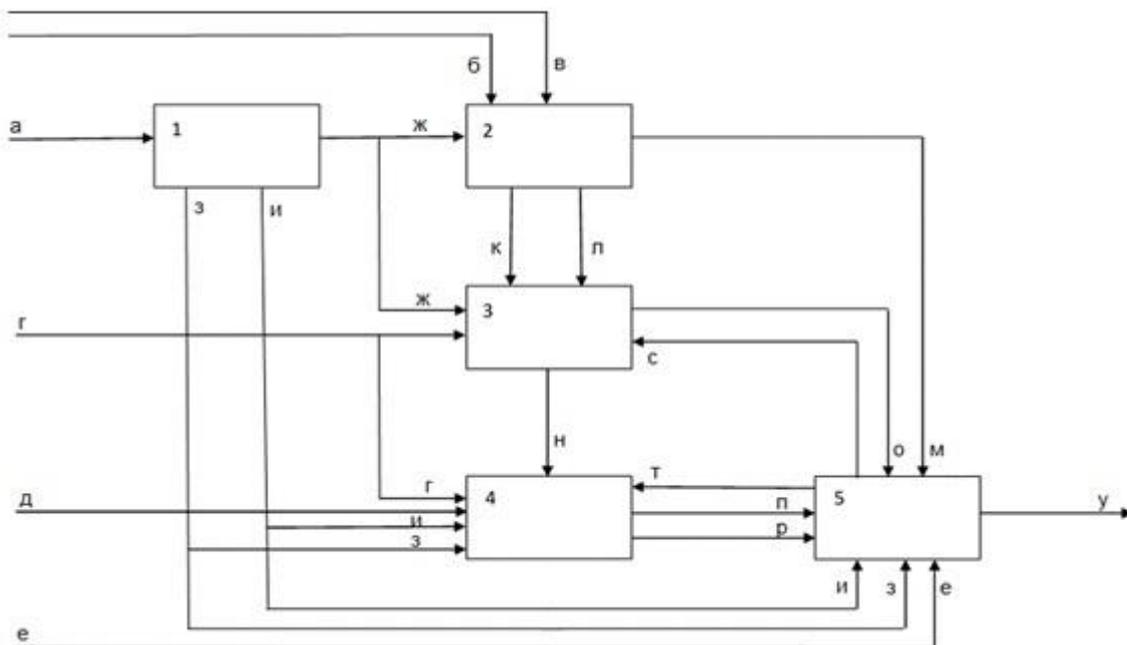


Рис. 1. Структурная модель системы отсеечения семантически нерелевантной и непертигентной текстовой информации

1. Булдакова А.А., Симонян Е.С., Кудрявцев А.Г. Выбор прототипа системы отсеечения семантически нерелевантной и непертигентной текстовой информации, Тезисы докладов IV Международной молодежной научной конференции ФТИ-2017 (2017).
2. Система ИРБИС64, Режим доступа: http://www.elnit.org/index.php?option=com_content&view=article&id=35&Itemid=108.
3. Гольдштейн С.Л., Кудрявцев А.Г., Разрешение проблемных ситуаций при поддержке систем, основанных на знаниях, ИД «Пироговъ» (2006).
4. Болбаков Л.Г., Теорема Байеса в когнитивной семантике образовательных информационных систем, Режим доступа: <http://cyberleninka.ru/article/n/teorema-bayesa-v-kognitivnoy-semantike-obrazovatelnyh-informatsionnyh-sistem>.
5. Лихтциндер Б.Я., Вероятностные оценки уровня знаний учащихся, Режим доступа: <http://ito.su/main.php?pid=26&fid=8028>.
6. Романов А.Н., Одинцов Б.Е., Советующие информационные системы в экономике, ЮНИТИ-ДАНА (2000).