

О МОДИФИКАЦИИ СИСТЕМЫ ОТСЕЧЕНИЯ СЕМАНТИЧЕСКИ НЕРЕЛЕВАНТНОЙ И НЕПЕРТИНЕНТНОЙ ТЕКСТОВОЙ ИНФОРМАЦИИ

Симонян Е.С. *, Кудрявцев А.Г.

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

*E-mail: Ses3004@mail.ru

ABOUT MODIFICATION OF EXCLUSION SYSTEM FOR SEMANTICALLY IRRELEVANT SPECIAL TEXT INFORMATION

Simonian E.S., Kudriavcev A.G.

Ural Federal University, Yekaterinburg, Russia

Annotation. A modification for working with accounting texts of the existing system exclusion irrelevant information is proposed. A structural model of the proposed system is presented. All changes made in the existing system are described and justified.

Известна система отсекающая семантически нерелевантную и непертинентную текстовую информацию [1-3, 4].

Нами предложена ее модификация для работы с бухгалтерскими текстами (Рис 1.).

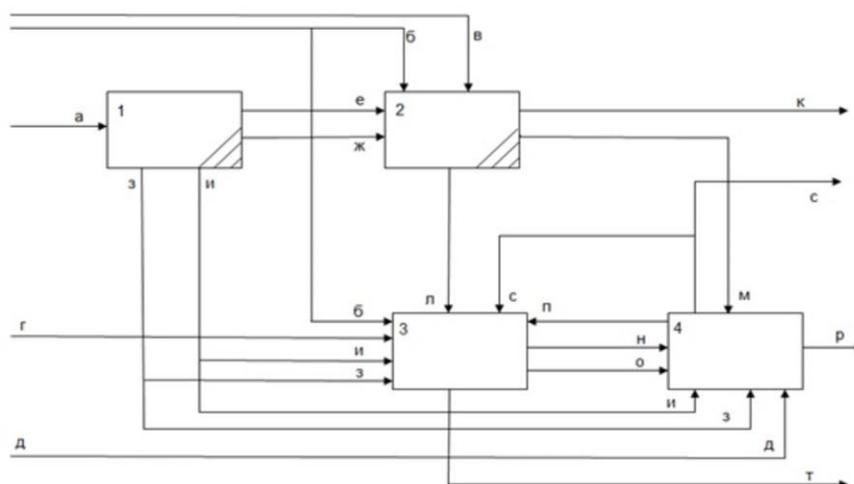


Рис 1. Структурная модель модифицированной системы отсекающей семантически нерелевантную и непертинентную текстовую информацию

(1 – препарататор (блок автоматического понимания текстов); 2 – блок взвешивания терминов; 3 – блок работы с семантическими связями; 4 – решатель; а – читаемый текст; б – взвешенный пользовательский словарь; в – пороги таксономии терминов, выбираемые пользователем; г – взвешенный пользовательский

список типовых семантических отношений; д – пороговые значения оценок Q_1 – Q_3 (см. ниже); е – неранжированный словарь читаемого текста; ж – коэффициенты ассоциативности для терминов словаря; з – таблица предложений читаемого текста; и – семантические структуры, обнаруженные в читаемом тексте; к – сведения о таксономии терминов для пользователя; л – взвешенный (по значимости терминов) словарь; м – оценка Q_1 терминологического соответствия читаемого текста пользователю; н – оценка Q_2 соответствия читаемого текста пользователю по семантическим связям; о – оценка Q_3 снятия семантической неопределенности [2,3] в читаемом тексте; п – стоп-сигнал для блока 3; р – сообщения о несоответствии; с – многоуровневый (по вложенности) реферат читаемого текста; т – графические иллюстрации реферата по уровням*); штриховка в правом нижнем углу блока означает его модификацию)

Из предложенного варианта изъят блок рубрицирования текстов, что связано как с отсутствием бухгалтерских текстов в ГПНТБ, так и с проблематичностью их рубрицирования как такового.

Для взвешивания терминов с помощью блока 2 предложено использовать сумму их коэффициентов ассоциативности с последующим таксономированием вместо вероятностной оценки задействования терминов при обнаружении рубрик.

Следует также отметить проблематичность извлечения терминов из бухгалтерских текстов традиционными методами, что приводит к необходимости модификации блока 1.

1. Савотченко С.Е. Современные аспекты повышения пертинентности результатов информационного поиска в глобальной сети [Электронный ресурс] / С.Е. Савотченко, Е.А. Проскурина. Режим доступа: <http://www.fundamental-research.ru/ru/article/view?id=34639>
2. Болбаков Л.Г. Теорема Байеса в когнитивной семантике образовательных информационных систем [Электронный ресурс] / Л.Г. Болбаков. Режим доступа: <http://cyberleninka.ru/article/n/teorema-bayesa-v-kognitivnoy-semantike-obrazovatelnyh-informatsionnyh-sistem>
3. Лихтциндер Б.Я. Вероятностные оценки уровня знаний учащихся [Электронный ресурс] / Б.Я. Лихтциндер, В.В. Пугин, Н.М. Татарина. Режим доступа: <http://ito.su/main.php?pid=26&fid=8028>
4. Булдакова А.А., Симонян Е.С., Кудрявцев А.Г. Выбор прототипа для системы отсеивания семантически нерелевантной и непертинентной специальной текстовой информации.