

**СОВРЕМЕННЫЕ ИНФОРМАЦИОННЫЕ
СИСТЕМЫ — ОСНОВА ЭФФЕКТИВНОЙ РАБОТЫ
БИБЛИОТЕК**

**MODERN INFORMATION SYSTEMS AS THE BASIS FOR
EFFECTIVE WORK OF LIBRARIES**

УДК 004.912-028.14:002.1:63

Д. П. Бабарико¹, А. Г. Буравкин², С. Ф. Липницкий², Л. В. Степура²

¹ГУ «Белорусская сельскохозяйственная библиотека им. И. С. Лупиновича»

Национальной академии наук Беларуси, Минск, Беларусь

²ГНУ «Объединённый институт проблем информатики

Национальной академии наук Беларуси», Минск, Беларусь

**СИСТЕМА АВТОМАТИЗИРОВАННОГО
РЕФЕРИРОВАНИЯ МНОГОЯЗЫЧНЫХ ЭЛЕКТРОННЫХ
МАССИВОВ НАУЧНО-ТЕХНИЧЕСКИХ ПУБЛИКАЦИЙ
ПО АГРАРНОЙ ТЕМАТИКЕ: ПРОГРАММНЫЕ И
ИНФОРМАЦИОННО-ЛИНГВИСТИЧЕСКИЕ СРЕДСТВА**

Аннотация. В докладе представлено описание программных и информационно-лингвистических средств системы автоматизированного реферирования многоязычных электронных массивов научно-технических публикаций по аграрной тематике, разрабатываемой в ОИПИ НАН Беларуси. Система предназначена для накопления и семантического сжатия текстовых документов по аграрной тематике на русском, английском и белорусском языках.

Ключевые слова: информационная система, автоматизированное реферирование, аналитическая обработка.

Для цитирования. Система автоматизированного реферирования многоязычных электронных массивов научно-технических публикаций по аграрной тематике: программные и информационно-лингвистические средства / Д. П. Бабарико [и др.] // Библиотеки в информационном обществе: сохранение традиций и развитие новых технологий. Тема 2018 года – «Научная библиотека как центр культурно-информационного пространства»: докл. III Междунар. науч. конф., Минск, 6–7 дек. 2018 г. / Белорус. с.-х. б-ка им. И. С. Лупиновича Нац. акад. наук Беларуси; редкол.: В. Н. Гердий [и др.]. – Минск, 2018. – С. 195–201.

D. P. Babaryka¹, A. G. Burawkin², S. F. Lipnicki², L. V. Stepura²

¹ State Institution «I. S. Lupinovich Belarus Agricultural Library» of the National Academy of Sciences of Belarus, Minsk, Belarus

² State Scientific Institution «The United Institute of Informatics Problems of the National Academy of Sciences of Belarus», Minsk, Belarus

SYSTEM OF AUTOMATED SUMMARIZATION OF MULTILINGUAL ELECTRONIC MASS OF SCIENTIFIC AND TECHNICAL PUBLICATIONS ON AGRARIAN SUBJECTS: SOFTWARE AND INFORMATION-LINGUISTIC MEANS

Abstract. The report presents a description of software and information-linguistic means of the automated summarization system of multilingual electronic mass of scientific and technical publications on agrarian topics developed at the United Institute of Informatics Problems of the National Academy of Sciences of Belarus. The system is designed for the accumulation and semantic compression of text documents on agrarian topics in Russian, English and Belarusian languages.

Keywords: information system, automated summarization, analytical processing.

For citation. Babaryka D. P., Burawkin A. G., Lipnicki S. F., Stepura L. V. System of automated summarization of multilingual electronic mass of scientific and technical publications on agrarian subjects: software and information-linguistic means. Libraries in the information society: preservation of traditions and development of new technologies. Theme 2018 – «A scientific library as a center for cultural and information space» : proceedings of the III international scientific conference, Minsk, 6–7 December 2018. Minsk, 2018, pp. 195–201 (in Russian).

Введение. В государственном научном учреждении «Объединённый институт проблем информатики» Национальной академии наук разработана система автоматизированного реферирования, которая предназначена для анализа многоязычных электронных массивов научно-технических публикаций по аграрной тематике, а также для создания рефератов статей и кратких сообщений из этих массивов. Данная разработка будет введена в эксплуатацию в государственном учреждении «Белорусская сельскохозяйственная библиотека им. И. С. Лупиновича» НАН Беларуси (БелСХБ) и обеспечит выполнение следующих функций:

- создание и актуализацию корпусов текстов по различным научным тематическим направлениям на русском, английском и белорусском языках;

- вычисление информативности слов и предложений в текстовых документах на русском, английском и белорусском языках;

- сжатие научно-технической информации и автоматизированное формирование рефератов текстовых документов в соответствии с заданными параметрами.

В докладе представлено программное и информационно-лингвистическое обеспечение системы автоматизированного реферирования.

Подготовка входных данных. БелСХБ является национальным отраслевым информационным центром в области аграрных наук. На подготовительном этапе из фонда библиотеки были отобраны документы-эталоны и документы на бумажных носителях, предназначенные для преобразования в электронную форму путём ретроспективной конверсии. В разработанной системе процессы реферирования текстовых документов основаны на использовании тематических и динамических корпусов текстов. Задачи их создания и использования решаются в рамках специального раздела языкознания — корпусной лингвистики. Под корпусом текстов понимают совокупность текстов, накопленных и размеченных по определенным принципам в зависимости от назначения.

В соответствии с критериями отбора (аграрная тематика) сотрудниками БелСХБ отобраны документы на русском, белорусском и английском языках. На их основе были сформированы тематические корпуса по аграрной тематике.

Информационно-лингвистическое обеспечение системы автоматизированного реферирования. Информационно-лингвистическое обеспечение системы автоматизированного реферирования состоит из специализированных баз данных и базы знаний:

- база данных с корпусами текстов;
- база данных для хранения результатов реферирования;

— база знаний, включающая специальные лингвистические словари (словарь словоформ, словарь словоизменительных парадигм, словарь синонимичных словоформ).

Для работы с тематическими и полными корпусами текстов на различных языках по заданным тематикам, а также для работы с лингвистическими словарями разработана подсистема актуализации лингвистической базы данных.

При создании корпусов текстов вводятся входные тексты и помещаются в базы корпусов текстов, релевантных этим текстам.

В подсистеме актуализации лингвистической базы данных (см. Рисунок 1) созданы средства для работы с тематическими и полными корпусами текстов на русском, белорусском и английском языках по различным тематикам.

В системе автоматизированного реферирования создаются лингвистические словари, используемые при индексировании текстовых документов и запросов пользователей (отдельно для каждого языка).

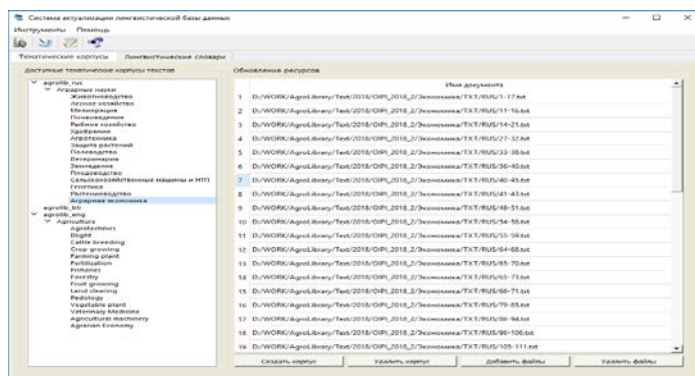


Рисунок 1 – Вкладка Тематические корпуса подсистемы актуализации лингвистической базы данных

При создании и актуализации словаря словоформ на вход поступает полный корпус текстов. На выходе получаем записи словаря (см. Рисунок 2), где каждой словоформе поставлены в соответствие:

- частота в полном корпусе текстов;
- частоты в тематических корпусах текстов;

– номер (код) парадигмы.

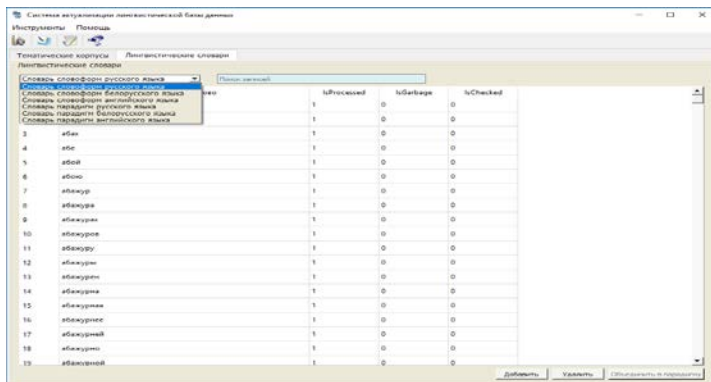


Рисунок 2 – Вкладка Лингвистические словари подсистемы актуализации лингвистической базы данных

Программное обеспечение системы автоматизированного реферирования. Подсистема синтеза рефератов многоязычных электронных массивов научно-технических публикаций реализуется в двух видах: в виде десктопной версии, а также в качестве веб-проекта, что предусматривает взаимодействие участников научно-исследовательских работ на основе удалённого доступа к интегрированным ресурсам.

С помощью данного инструментария (см. Рисунок 3) можно осуществлять следующие операции:

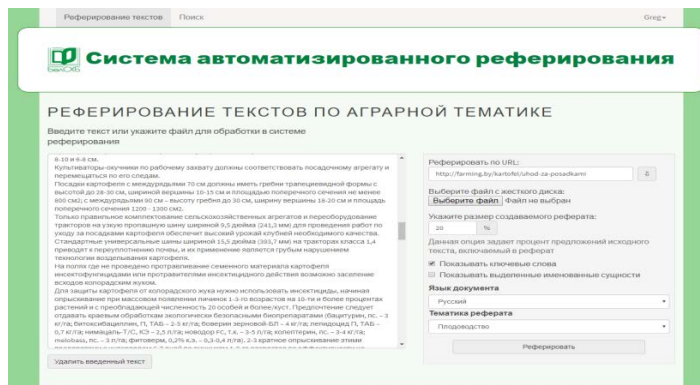


Рисунок 3 – Веб-страница подсистемы синтеза рефератов

- реферирование текстовых документов из различных информационных источников (жесткий диск компьютера пользователя, Интернет, локальная сеть);
- редактирование результатов реферирования текстовых документов;
- визуализация результатов реферирования документов (см. Рисунок 4):

Язык документа:Русский(установлен вручную)	
Реферат документа	
Уход за картофелем на лёгких почвах. Применяются ротационные бороны и рыхлители, а также культиваторы с пассивными рабочими органами (окучники, долота, стрелчатые лапы). Уход за картофелем на тяжёлых почвах: в случае применения фрезерных культиваторов уход за картофелем сводится к одной процедуре - формирование гребней высотой 20 – 25 см. Основной уход за картофелем включает следующие виды работ: – борьба с сорняками; – окучивание с целью формирования гребней; – поддержание почвы в рыхлом состоянии. При выпадении обильных осадков междурядья рыхлят ярусными окучниками, чтобы не было трещин.	
Ключевые слова	Информативность
ярусными	8.33
насыпается	8.33
стрелчатые	3.45
рыхлители	2.38
рыхлост	2.04
долота	2.00
бороны	1.92
окучниками	1.54
боронования	1.28
обрабатываются	1.20

Рисунок 4 – Результат реферирования

В работах [1, 2] представлены алгоритмы реферирования и выявления контекста информативных предложений.

В системе реализована возможность сохранения созданных рефератов в базу данных с дополнительной информацией, такой как выделенные ключевые слова и именованные сущности.

Заключение. В результате работ по данному проекту:

- были созданы алгоритмы, программные и информационно-лингвистические средства системы автоматизированного реферирования многоязычных электронных массивов научно-технических публикаций на белорусском, русском и английском языках;

- планируется введение в эксплуатацию указанной системы автоматизированного реферирования в государственном учреждении «Белорусская сельскохозяйственная библиотека им. И. С. Лупиновича» НАН Беларуси.

Список использованных источников:

1. Липницкий, С. Ф. Поиск и реферирование текстовой информации в многоязычной среде / С. Ф. Липницкий, А. А. Мамчич, Л. В. Степура // Открытые семантические технологии проектирования интеллектуальных систем = Open Semantic Technologies for Intelligent Systems (OSTIS–2013) : материалы III Междунар. науч.-техн. конф. (Минск, 21–23 февр. 2013 г.) / Белорус. гос. ун-т информатики и радиоэлектроники ; редкол.: В. В. Голенков (отв. ред.) [и др.]. – Минск, 2013. – С. 229–232.
2. Степура, Л. В. Автоматизированное реферирование многоязычных электронных массивов научно-технических публикаций / Л. В. Степура // Труды Конгресса по интеллектуальным системам и информационным технологиям «IS&IT'18», Дивноморское, Россия, 2–9 сентября 2018 г. / Юж. федер. ун-т [и др.]. – Таганрог, 2018. – Т. 1. – С. 120–124.

Поступила в редакцию 05.10.2018
Received 05.10.2018