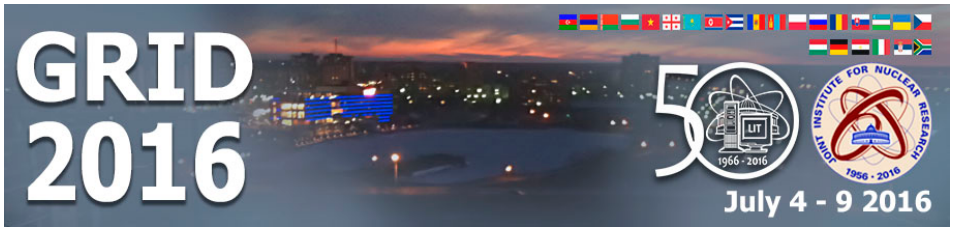

Contribution ID: **61** Type: **Sectional reports**

Создание, поддержка и развитие модели интерпретации смыслов

Thursday 7 July 2016 16:15 (15 minutes)

В мире наблюдается постоянный и устойчивый рост количества научных рецензируемых журналов и публикуемых в них статей. [1,2]. По оценкам некоторых авторов в мире издается порядка 1 млн. статей в год, что соответствует ежедневному выпуску порядка 2700 публикаций [2]. Хотя значительная часть такого рода ресурсов находится в электронных фондах (ЭФ), тщательно изучать такие информационные массивы для ученых и исследователей традиционными способами становится все труднее. Но смысловой поиск –это необходимый этап, предшествующий генерации нового знания в научной среде, и порядка 60% времени ученый тратит именно на поиск научной информации [3]. Научные статьи представляют собой тексты на естественном языке, отвечающие определенным требованиям к структуре и содержанию: однозначность, логика от посыла к следствию, явная целевая установка статьи, ясность и точность. Несмотря на шаблон, в них есть неопределенность, связанная с неоднозначностью интерпретации читателем.
В работе излагается технология извлечения смысла из научных текстов на основе модели интерпретации (донесения до определенной аудитории) смысла, состоящей из следующих компонентов: 1) выделение смысла из научной статьи (выжимка, словарь, семантическая модель (слова и связи)); 2) формирование логико-семантической модели (сети); 3) вопросно-ответный параметрический навигатор. Такая модель используется для структурирования информационных фондов на основе каталожной службы, представляющей собой множество логико-семантической сетей (ЛСС) [4]. Эффектом является сокращение времени на изучение фонда за счет повышения уровня понимания данной статьи.
Организация такой службы является трудоемкой задачей и требует от эксперта (аналитика ЭФ) выполнения большого объема работы для анализируемого информационного ресурса. В статье исследуется возможность автоматического анализа научных текстов на предмет выявления основных и скрытых смыслов с использованием набора техник интеллектуального анализа текстов (Text Mining). Результаты анализа будут использоваться экспертом для построения и поддержки набора ЛСС. На основе описанной модели создан прототип, частично реализующий функционал АРМ аналитика ЭФ и вопросно-ответный навигатор.

[1] Т. Н. Домнина, О. А. Хачко (ВИНИТИ РАН).Научные журналы: количество, темпы роста//Информационное обеспечение науки: новые технологии, Сб. науч. тр. / Каленов Н.Е., Цветкова В.А. (ред.). - М.: БЕН РАН, 2015.-с.83-96. [Электронный ресурс] —URL: http://www.benran.ru/SEM/Sb_15/sbornik/83.pdf
[2] Галина Якшонок. Эффективный поиск и анализ научно-исследовательской информации в Sci-Verse: Scopus, Hub, ScienceDirect//МГИМО, 2012. [Электронный ресурс]. URL: http://mgimo.ru/files2/y03_2012/220642/MGIMO_Mar 2012.ppt
[3] Коэн Э. Сравнительный анализ научных публикаций российских и зарубежных ученых. - 2006 //«Эльзевир» [Электронный ресурс] —URL: http://www.elsevier.ru/about/articles/?id=792
[4] Добрынин В.Н., Филозова И.А. Семантический поиск в научных электронных библиотеках//Информатизация образования и науки № 2(22)/2014. - с.110-110

Author: FILOZOVA, Irina (JINR)

Co-author: DOBRYNIN, Vladimir (Dubna State University, Dubna, Russian Federation)

Presenter: FILOZOVA, Irina (JINR)

Session Classification: 10. Databases, Distributed Storage systems, Big data Analytics

Track Classification: 1. Technologies, architectures, models of distributed computing systems