

УДК 004.912

Визуализация неструктурированных текстов

Д.С.Володин, К.А. Гудков, А.А. Даниляк, П.М. Дубов, А.А. Журавлёв, И.И. Ивашнёв, С.Э.о. Исалы, И.А. Карацапова, Н.С.Карацапова, Н.И. Кривчанский, Д.В. Кудрявцев, А.В. Куликов, Е.В. Луничкин, П.В.Матавина, М.М. Сандрикова, А. Саутин, А.А. Смирнов, А.В. Чупахина, О.В. Чупахина, А.Ф. Халяпов

Московский физико-технический институт (государственный университет)
АВВУУ

Maria_Sa@abbyy.com, Alexander_D@abbyy.com, Nadezda_K@abbyy.com, Polina_Ma@abbyy.com

В современном мире ежегодно появляются сотни терабайт неструктурированной текстовой информации, которую необходимо обрабатывать машинными методами. Например, часто возникает потребность в выделении различных фактов из текстов и представлении их в наглядном виде.

Одним из решений такой задачи является двухступенчатая система, которая использовалась в нашем проекте. На первом этапе исходный текст преобразовывался в структурированный при помощи технологии извлечения информации на основе синтактико-семантического анализа текстов на естественном языке АВВУУ Compreno. Затем полученные данные были визуализированы с использованием различных open-source решений.

Для демонстрации результатов были выбраны приключенческие романы Жюль Верна.

Литература

1. *Starostin A. S., Smurov I.M., Stepanova M.E.* A Production System for Information Extraction Based on Complete Syntactic-Semantic Analysis // *Computational Linguistics and Intellectual Technologies: papers from the Annual conference “Dialogue”* — 2014
2. *Hogenboom F., Frasinca F., Kaymak U., De Jong F.* An overview of event extraction from text // *Workshop on Detection, Representation, and Exploitation of Events in the Semantic Web (DeRiVE 2011) / Tenth International Semantic Web Conference (ISWC 2011)*, Vol. 779, pp. 48–57. — 2011
3. OWL Web Ontology Language Overview, <http://www.w3.org/TR/2004/REC-owl-features-20040210> — 2004