

ХАЛҚАРО ИЛМИЙ АНЖУМАН  
МЕЖДУНАРОДНАЯ НАУЧНАЯ КОНФЕРЕНЦИЯ  
INTERNATIONAL SCIENTIFIC CONFERENCE

# INNOVATION -2014

Илмий мақолалар тўплами  
Сборник научных статей  
Proceedings of the Conference



Ташкент

23-24 октябрия

УЎК:021.13

КБК: 30

И-58

Innovation – 2014. Илмий мақолалар тўплами. – Тошкент: Янги аср авлоди,  
2014. - 348 б

ISBN 978-9943-27-391-7

ISBN 978-9943-27-391-7

INNOVATION – 2014. Илмий мақолалар тўплами. “Янги аср авлоди”, 2014 й.

## **AUTOMATION OF LINGUISTIC EXAMINATION OF ELECTRONIC DOCUMENTS**

**M. Gvozdenko**

The report is devoted to topical questions of creation of methods; programs and information systems for identifying the author of the text and, in particular, the electronic documents. A list of modern programs for determination of the author's text and their description are given.

### **АВТОМАТИЗАЦИЯ ЛИНГВИСТИЧЕСКОЙ ЭКСПЕРТИЗЫ ЭЛЕКТРОННЫХ ДОКУМЕНТОВ**

**Гвозденко М.В.** (НЮУ им.Я.Мудрого, г.Харьков, Украина)

Стремительное развитие сетевых технологий привело к значительному росту количества электронных текстовых документов. Одновременно с этим возросло и количество нарушений интеллектуальных прав авторов текстовых произведений, возникли проблемы массового распространения конфиденциальной информации, информации, что нарушает достоинство и деловую репутацию отдельных лиц и учреждений. Научные статьи, диссертационные работы, учебники, произведения художественной литературы и т.д. бесконтрольно распространяются в сети Интернет, иногда без всяких редакторских изменений, иногда после незначительного редактирования. И если прямое копирование, так называемый Copy-Paste, не требует проведения экспертного исследования для выявления автора текста, то текст, отредактированный огромным штатом достаточно осведомленных рерайтеров, требует очень тщательного и достоверного экспертного исследования. Ведь на сегодняшний день рерайтеры используют широкий арсенал средств преобразования уникального текста в тиражирование нескольких текстов: перевод уникального текста на другой язык и редактирование полученного перевода, использование синонимов (даже с использованием специальных программ - синонимайзеров), замена прямой речи на косвенную, добавление в уникальный текст или удаления из него слов и словосочетаний, которые не несут смысловой нагрузки, использование нескольких уникальных текстов для формирования нового текста и др.

Соответственно, значительно возросла потребность в выполнении лингвистических экспертиз. Поскольку повышение качества и скорости выполнения экспертных исследований в значительной степени зависит от использования средств информационных технологий, автоматизация лингвистической экспертизы электронных документов на сегодняшний день приобретает большую актуальность.

Основной проблемой установления авторства является именно определение личных признаков, причем эти признаки должны отвечать нескольким условиям: признак должна отражать такие особенности текста,

которые автор создает подсознательно, они должны сохранять постоянное значение для одного автора и иметь существенно разные значения для разных авторов. Чем больший объем текста предлагается для экспертного исследования, тем проще обнаружить стабильные частоты некоторых языковых элементов, присущих именно данному автору. В разработке компьютерных программ определения авторства используются формальные методы идентификации автора текста, которые основаны на сравнении вычисляемых характеристик текстов, таких как подсчет частоты и природы лексических, орфографических, синтаксических и грамматических ошибок; исследования стилистических факторов письменного языка (длина слов, длина предложений; количество складов, приставок и суффиксов на 100 слов); подсчет процента имеющих в тексте частей речи: соотношение глаголов к прилагательным, глаголов - к существительным и т. п., а также показатель TTR (TypeTokenRatio) - представление в форме десятичной дроби соотношения количества различных слов с общим количеством слов в тексте.

В докладе рассмотрены формальные методы, которые положены в основу разработки и функционирования компьютерных программ проведения лингвистической экспертизы: рассмотрены метод сравнения гистограмм частот имеющихся слов разной длины, 'наивный байесовский' (НБА) метод, метод распределения параметра, сравнение количества новых слов в спорном тексте, метод относительной энтропии, метод устойчивости частот, индекс Флеша, FOG-индекс, подход Колтарда, лингво-статистический анализ неполнозначной лексики, распознавание автора текста с использованием цепей А.А. Маркова, атрибуция Минимальной Условной Сложности Сжатия (МУСС - атрибуция), аппарат машины опорных векторов (SupportVectorMachines, SVM). Доказана преимущество аппарата нейронных сетей по сравнению со статистическими методами. В докладе сделан обзор возможностей современных программ лингвистической экспертизы и их классификация. К сожалению, ни одна из существующих программ не дает стабильного достоверного результата, особенно на небольших фрагментах текста.

## **TECHNOLOGICAL ASPECTS OF MULTISERVICE NETWORK**

**G. Sherboboeva**

The article shows a simplified model of care enabling analysis of the basic technological aspects of construction and development of multiservice network.

## **ТЕХНОЛОГИЧЕСКИЕ АСПЕКТЫ СОЗДАНИЯ МУЛЬТИСЕРВИСНЫХ СЕТЕЙ**

**Шербобоева Г.Б.**

**Editorial Board:** A. Kadirov (Chairman), H. Bakirova, I. Bibik, A. Kadirova, Sh. Madaminov, K. Matyakubov, M. Sultanova

**Редацион коллегия:** Қодиров А.А. (раис), Бакирова Х.Б., Бибиқ И.П., Қодирова А.А., Мадаминов Ш.А, Матякубов К.Р., Султанова М.М.

**Редационнаа коллегия:** Кадыров А.А. (председатель), Бакирова Х.Б., Бибиқ И.П., Кадырова А.А., Мадаминов Ш.А, Матякубов К.Р., Султанова М.М.

**International Scientific and Practical Conference  
«INNOVATION-2014»**

**«ИННОВАЦИЯ-2014»  
Халқаро илмий-амалий Анжумани**

**Международная научно-практическая Конференция  
«ИННОВАЦИЯ-2014»**

*Proceedings of the Conference  
Илмий мақолалар тўплами  
Сборник научных статей*

*Компьютерная вёрстка:  
Пулатов У.У., Саиджанов Р.К.*