

основні концептуальні аспекти теорії нечіткої логіки: функції належності, лінгвістичні змінні, нечіткі логічні операції, методи нечіткої імплікації та нечіткої композиції. Системи нечіткого висновку призначені для реалізації процесу обробки інформації і слугують концептуальним базисом усієї сучасної нечіткої логіки. Механізм або алгоритм висновку є важливою частиною базової архітектури систем нечіткого висновку.

#### СПИСОК ЛІТЕРАТУРИ

1. Ольховська О. Л. Економіко-математична модель діагностики банкрутства страхової компанії на основі нечіткої логіки / О. Л. Ольховська // *Моделювання та інформаційні системи в економіці*. – К. : КНЕУ, 2010. – Вип. 81. – С. 59–74.
2. Ольховська О. Л. Оптимізація нечіткої моделі оцінки фінансового стану страхової компанії із застосуванням градієнтного підходу / О. Л. Ольховська // *Євроінтеграція економіки України: виклики та рішення: матеріали Міжнародної науково-практичної конференції, 20–21 лютого 2015 р.* – Національний гірничий університет. – Дніпропетровськ: Видавничий дім «Гельветика», 2015. – С. 259–262.
3. Ольховська О. Л. Моделювання фінансово стану страхової компанії: монографія / О. Л. Ольховська, А. В. Матвійчук. – Краматорськ : ДДМА, 2015. – 128 с.
4. Матвійчук А. В. Моделювання економічних процесів із застосуванням методів нечіткої логіки : монографія / А. В. Матвійчук. – К. : КНЕУ, 2007. – 264 с. – ISBN 966-574-966-8.

## РАЗРАБОТКА ИНФОРМАЦИОННОЙ СИСТЕМЫ МОРФОЛОГИЧЕСКОГО АНАЛИЗА ТЕКСТОВ

Птущук В. Ю., Вычужанин В. В.  
ОНМУ, г. Одесса

Компьютерный лингвистический анализ текстов – перспективная и быстро развивающаяся область искусственного интеллекта. Одна из ключевых задач компьютерного лингвистического анализа состоит в построении такого структурированного представления текста, к которому можно применять методы и алгоритмы решения прикладных задач. Существует значительное количество разновидностей методов как синтаксического, так и семантического анализа, которые основаны на разных моделях синтаксической структуры предложения и разном понимании семантики.

На практике, активное распространение получают методы, позволяющие анализировать текстовую информацию. За счет того, что текстовые документы представляют собой большой массив информации, то использование простых алгоритмов не является эффективным решением задачи анализа данных. Использование методов и алгоритмов анализа, которые входят в состав Data Mining, может существенно повысить эффективность в обработке текстовой информации и дальнейшего использования полученных результатов. При этом различают подходы, используемые в информационно-поисковых системах и системах анализа и обработки данных

(Mining). Методы Text Mining предназначены для проведения морфологического анализа текстов, которые позволяют анализировать большое количество текстовых документов за небольшой промежуток времени и входят в состав концепций, которые разрабатываются в рамках области искусственного интеллекта. Задачи Text Mining — выбирать из текстов наиболее важную и значимую информацию для пользователей [1]. К основным элементам Text Mining относятся [2]:

- классификация (classification, categorization);
- кластеризация (clustering);
- извлечение фактов, понятий (feature extraction);
- реферирование (summarization);
- ответ на запросы (question answering);
- тематическое индексирование (thematic indexing);
- поиск, по ключевым словам, (keyword searching);
- построение графика стилистической особенности;
- выявление авторства текстов.

При классификации текстов используются статистические корреляции для отнесения документов к определенным категориям. Задача классификации — это классическая задача распознавания, где по некоторой контрольной выборке система определяет категорию нового объекта. Особенность классификации в рамках концепции Text Mining заключается в том, что количество объектов и их атрибутов может быть очень большим, поэтому должны быть предусмотрены механизмы оптимизации этого процесса [3].

Разработка информационной системы морфологического анализа текстов, рис. 1, включает в себя этапы формирования файла для дальнейшего анализа, использование аналитических методов Text Mining, а также применение методов оценки полученных в результате анализа результатов. В основе аналитического метода применяется факторный анализ, а также алгоритм pLSA, которые позволяют провести сравнительный анализ отдельно взятых термов или сопоставить два документа между собой. На практике это реализуется за счет использования матричного индексирования и сопоставления результатов друг другу. Матрица, которая содержит индексные значения становится отображением коллекции документов в векторном гиперпространстве, что позволяет сопоставлять вектора между собой в качестве анализа.

Использование алгоритмов группы LSA подразумевает отображение документа в латентное семантическое пространство, которое несет в себе смысловую нагрузку. Целью такого отображения является выявление латентной сингулярной связи между термами, за счет использования сингулярного разложения матриц (SVD). Оценка схожести термов формируется по близости расположения точек латентного семантического пространства. Комплексное использование перечисленных алгоритмов, позволяет классифицировать документы по уникальному стилю написания, однако, для работы такой проверки, необходимо большое количество материалов для обуче-

ния. В тоже время, анализировать документ на количество термов, а также вывод статистики по количеству использования термов, их стилистической привязки, совместном использовании с другими термами не требует.

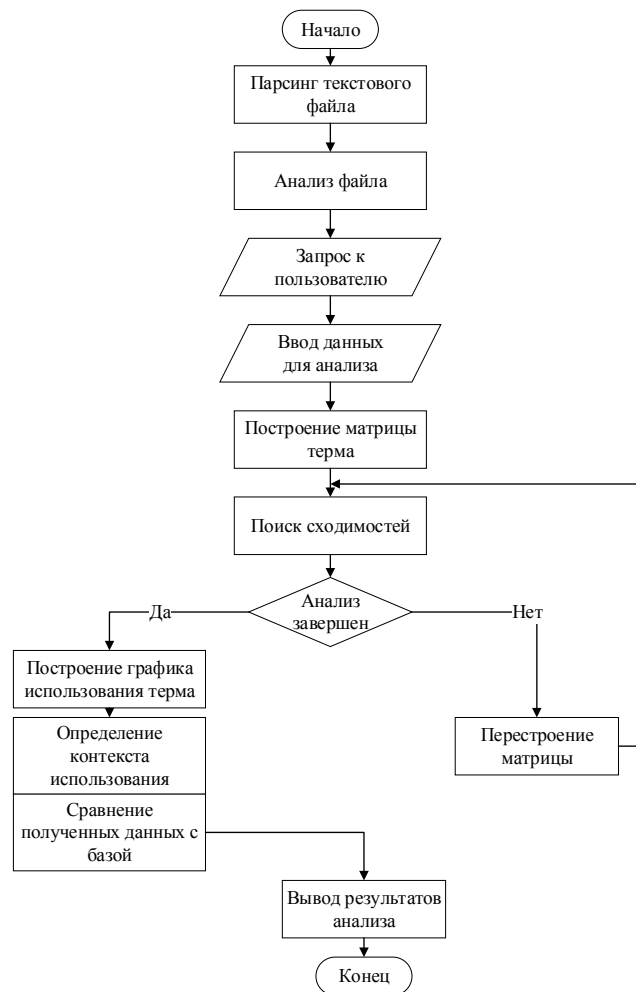


Рисунок 1 – Блок-схема работы информационной системы

## ВЫВОДЫ

Результатом проведенной работы является разработка информационной системы морфологического анализа текстового файла. Информационная система позволяет выводить статистический анализ использования терма, его периодичность, связанность с другими термами, а также анализировать персональную стилистику автора для дальнейшей классификации текстового файла.

## СПИСОК ЛИТЕРАТУРЫ

1. Berry M. W., Kogan J. *Text Mining. Applications and Theory*. – Wiley. – 2010. – 207 p.
2. Ландэ Д. В. *Интернетика. Навигация в сложных сетях: модели и алгоритмы* / Д. В. Ландэ, А. А. Снарский, И. В. Безсуднов. – М. : Либроком, 2009. – 264 с.
3. Lomakina L. S. *Hierarchical Clustering of Text Documents* / L. S. Lomakina, V. B. Rodionov, A. S. Surkova // *Automation and Remote Control*. – 2014. – Vol. 75. – N. 7. – P. 1309–1315.