

## ИНТЕЛЛЕКТУАЛЬНЫЕ АГЕНТЫ ПРИ ПОИСКЕ В РАСПРЕДЕЛЕННЫХ ИНФОРМАЦИОННЫХ СТРУКТУРАХ

*В.В. Лаврентьев, Ю.В. Савицкий*

Брестский государственный технический университет  
кафедра Интеллектуальных информационных технологий  
e-mail: stropa@gmail.com +375 29 386-88-80

Полнотекстовый поиск в распределенных информационных структурах (базы знаний распределенные по ресурсам локальной сети, в Интернет, в распределенных базах данных, хранилищах данных) является одной из сфер применения интеллектуальных алгоритмов [1] основанных на исследованиях в области искусственного интеллекта.

С ростом и развитием корпоративных сетей, включая использование возможностей Интернет, базы знаний выходят за рамки сбора локальной информации. Полнотекстовый поиск без предварительной индексации, вне зависимости от технологичности реализации, не позволяет достичь адекватного времени отклика (временного промежутка запрос-ответ) [2] при работе даже в условиях современных предприятий малого и среднего бизнеса (в зависимости от специфики деятельности). Индексация данных является наиболее распространенным на данный момент [1, 3] методом получения малого времени отклика, при поиске на больших выборках данных. Однако данный метод не отвечает требованиям актуальности информации и ее местонахождения [3, 4], а значит не соответствует приоритетным условиям поставленной задачи.

Замена предварительной индексации данных на работу интеллектуальных (рациональных) агентов должна привести к удовлетворительному компромиссу между скоростью обработки информации при поиске и актуальностью полученной выборки. Это является ключевой задачей на данном этапе. Релевантность результирующей выборки используется как вторичный параметр оценки результата [5] и выступает скорее в роли контрольной характеристики адекватности функционирования поисковой системы в целом.

Суть использования интеллектуальных агентов заключается в разработке автономно функционирующего модуля [1], в задачи которого входит осуществление поиска запрошенной информации непосредственно в рассматриваемой предметной области. Это является одним из ключевых отличий от традиционных систем поиска с использованием интеллектуальных агентов [1, 6], где функции агентов заключаются в сборе информации для целей построения индексной базы данных.

Преимущество метода заключается в возможности использования нескольких агентов одновременно для целей одного поискового запроса.

В процессе работы системы ведется статистика и анализ параметров каждого запроса, таких как: время отклика, относительная релевантность результата, объем результирующей выборки. Количество активируемых агентов система выбирает автоматически [6] на основе предыдущего опыта. Изначально генерируется несколько базовых агентов, каждый из которых получает уникальные стартовые характеристики [6] (метод обхода информационной базы, матрицу приоритетов «клиент – тип информации» и другие). На их основе генерируются действующие экземпляры (клоны). Все экземпляры на основе одного базового объекта работают на развитие общих параметров (накапливают общий опыт). Агенты, регрессирующие и остановившиеся в развитии, продолжают существовать заранее заданный промежуток времени (количество активированных экземпляров). Таким образом, предполагается достичь большей скорости получения интеллектуальных агентов, действующих рационально. Цель механизма – прийти к времени, сопоставимому со временем отклика при индексации, за время, сравнимое со временем полной предварительной индексации данных. Оценка временных характеристик производится на одной и той же информационной структуре, упрощенной для целей полнотекстового поиска доступными существующими средствами поиска с индексацией.

На текущем этапе разработки установлено постоянное число клиентов (пользователей) системы. В перспективе данный параметр будет введен в разряд подлежащих обучению, т.к. основная идея системы – максимальная автономность с начала работы в рамках поставленной задачи [1] поиска информации, включая поиск информационных ресурсов, как внутренних, так и внешних (автоматическое расширение базового информационного пространства в пределах информационной потребности клиентов системы).

Другим преимуществом данного метода является возможность наделить агента способностями к интеллектуальному анализу информации [1]. Здесь реализуется попытка найти компромисс между достаточно медленным полнотекстовым поиском и неполной индексацией. Использование алгоритмов интеллектуального анализа информации на основе разработок в области построения семантических сетей [7], позволяет получить меньшее время анализа объекта (документа, сущности) без потери информации или с минимальным уровнем потери. Для повышения качества поиска, при разработке поискового модуля интеллектуального агента, используются алгоритмы на основе сетей Хемминга и других подобных методов [8, 9],

позволяющих минимизировать ошибку поискового запроса и учесть особенности естественного языка.

### ЛИТЕРАТУРА

1. Рассел С., Норвиг П. Искусственный интеллект – Вильямс, 2007. 1410 с.
2. Кормалев Д. А., Куршев Е. П., Осипов Г. С., Сулейманова Е. А., Трофимов И. В.: Препринт // Методы поиска и анализа информации. Автоматическое извлечение данных.– Переславль-Залесский: ИПСРАН, 2003.
3. Junghoo Cho, Hector Garcia-Molina. The Evolution of the Web and implications for an Incremental Crawler. – Department of Computer Science Stanford, 2000.
4. Колисниченко Д. Н. Поисковые системы и продвижение сайтов в Интернете. — М.: Диалектика, 2007. 272 с.
5. Беляев Д.В. Ассоциативная модель смысловых контекстов и ее применение в задаче уточнения поисковых запросов. // Электронный журнал "Труды МАИ", 2005, N18 [http://www.mai.ru/projects/mai\\_works/articles/num18/article9/author.htm](http://www.mai.ru/projects/mai_works/articles/num18/article9/author.htm)
6. Козлов Е. Б., Метелкин А. В., Хорошевский В. Ф. Мультиагентная система поиска информации в Интернет // Труды седьмой национальной конференции по искусственному интеллекту с международным участием КИИ-2000. – М.: Физматлит, 2000. 840 с.
7. Sowa J. F. Semantic networks. (Revised and extended version of an article originally written for the Encyclopedia of Artificial Intelligence, edited by Stuart C. Shapiro, Wiley, 1987, second edition, 1992) <http://www.jfsowa.com/pubs/semnet.htm>
8. Бойцов Л.М. Использование хеширования по сигнатуре для поиска по сходству. Прикладная математика и информатика. – М.: Издательство факультета ВМиК, МГУ 2000, № 7.
9. Блейхут Р. Теория и практика кодов, контролирующих ошибки – М.: Мир, 1986.