

# ИСПОЛЬЗОВАНИЕ ИНТЕЛЛЕКТУАЛЬНЫХ АГЕНТОВ ДЛЯ ПОИСКА ИНФОРМАЦИИ В ИНТЕРНЕТ

Козлов Д.Д., Смелянский Р.Л.

Московский Государственный Университет им. М.В. Ломоносова, факультет Вычислительной Математики и Кибернетики, Лаборатория Вычислительных Комплексов, 119899, Россия, Москва, ГСП-3, Воробьевы горы, МГУ, 2-й учебный корпус, ф-т ВМК, тел. 9394671, факс 9392596, e-mail: ddk@cs.msu.su, smel@cs.msu.su

## ABSTRACT

In this paper personal intelligent assistant for information retrieval is discussed. Most of discussion concerns intelligence feature and it's implementation via user and environment modeling.

## ВВЕДЕНИЕ

Поиск нужной информации в сети Интернет представляет собой существенную проблему в связи с огромным объемом доступной информации. Широко распространенные на сегодня средства поиска не способны обеспечить высокое качество поиска. Наиболее развитые возможности поиска предоставляют сегодня системы поиска по ключевым словам. Существенным недостатком таких систем является низкая точность выдаваемой информации. Развитие возможностей этих систем, направленное на повышение точности информации приводит к усложнению языка запросов этих систем. Кроме того, в каждой из систем имеется свой синтаксис языка запросов. В результате, большинство пользователей просто игнорируют развитые возможности систем поиска и используют только базовые возможности, приводящие к низкому качеству результатов поиска. Данная ситуация приводит к тому, что пользователь вынужден сам обрабатывать большое количество документов.

Для постоянной работы пользователя в Интернет (именно работы, с целью получения нужной информации, а не “свободного плавания”) характерна долговременная заинтересованность пользователя в информации по некоторым узким областям.

Поэтому целесообразно предоставить пользователю персонального информационного помощника (далее агента), который с одной стороны позволил бы автоматизировать задачу сбора и накопления тематической информации, а с другой бы повысил качество поиска в Интернет.

На наш взгляд такой информационный помощник должен обеспечивать следующие возможности:

- поиск информации по заданию пользователя однократный и постоянный;
- возможность автономного выполнения заданий человека;
- возможность доставки информации на компьютер пользователя;
- фильтрация постоянных потоков информации;
- предоставление вновь появляющейся в Internet информации, которая может заинтересовать пользователя, в связи с каким-нибудь из его запросов.

## ОБЩАЯ СТРУКТУРА ИНФОРМАЦИОННОГО ПОМОЩНИКА

Особый интерес представляет построение информационного помощника в виде интеллектуальной системы, так называемого *интеллектуального* агента. Интеллектуальность агента является степенью способности к рассуждению и обучаемости. Интеллектуаль-

ность подразумевает, как минимум, возможность задавать пользовательские предпочтения агенту и наличие у агента механизма рассуждения, чтобы действовать в соответствии с этими предпочтениями. Более высокий уровень интеллектуальности подразумевает наличие у агента модели пользовательских потребностей и механизма поиска способа их удовлетворения.

Исходя из заявленной задачи, можно выделить ряд основных свойств, которыми должен обладать такой агент.

- *Автономность* - агент должен выполнять большую часть своей работы автономно, не взаимодействуя с человеком или другими агентами.
- *Коммуникабельность* – агент должен уметь общаться с пользователем, получая от него задания и предоставляя результаты.
- *Адаптируемость и адаптивность поведения* – в ходе общения с пользователем агент должен уметь настраиваться (или, хотя бы быть настраиваемым) под привычки и методы работы конкретного пользователя.
- *Рациональность поведения* - агент своими действиями должен продвигаться к решению поставленной задачи и не делать действий, этому процессу препятствующих.
- *Восприимчивость* – агент, находясь в окружающей его информационной среде, должен воспринимать некоторым образом изменения окружающей среды и реагировать на изменения.
- *Проактивность* - агент не только должен формально выполнять поставленную задачу поиска, но и должен собирать при этом полезную для пользователя информацию, относящуюся к запросу пользователя.

Важно отметить, что агент не является информационно-поисковой системой, как, например, Yahoo, а является посредником между человеком и средствами поиска, доступными в Интернет, добавляя в то же время и собственную поисковую функциональность. Агент должен поддерживать накопление полезной для пользователя информации в своей базе данных.

Агент не заменяет имеющихся средств поиска информации, а взаимодействует с ними, используя их как средства воздействия с целью выполнения задачи и как средства получения информации об окружающей среде.

Исходя из вышеперечисленных свойств, можно предложить следующую организацию агента (см. рис. 1).

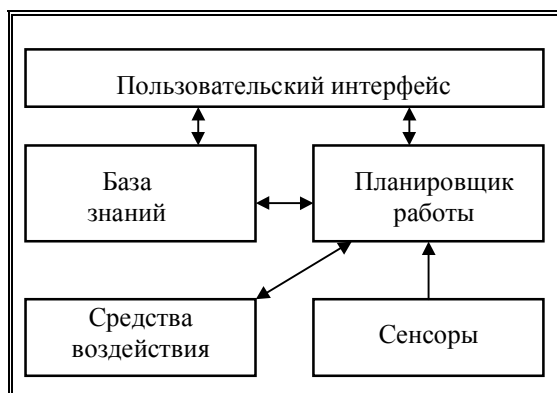


Рис. 1. Общая структура агента

Агент работает постоянно, выполняя несколько заданий пользователя одновременно. Диспетчеризацией действий агента направленных на работу с внешней средой и с пользователем выполняет *планировщик*.

- Планирование поиска.
- Осуществление поиска.
- Автоматическое (без участия пользователя) пополнение базы знаний.
- Взаимодействие с другими агентами.

База знаний предназначена для представления следующей информации:

- модели пользователя;
- модели окружающей среды (Internet);
- информации о доступных средствах воздействия на окружающую среду;
- информации о других агентах, с которыми можно взаимодействовать;
- модели предметной области.

Сенсоры и средства воздействия осуществляют связь агента с окружающей средой.

Пользовательский интерфейс служит для получения заданий от пользователя и предоставления пользователю результатов выполнения заданий. Учитывая то, что агент строит внутри себя модель пользовательских потребностей очень полезно предоставить пользователю графическое представление этой модели и возможность ее редактирования. Кроме того, пользовательский интерфейс служит для ручного пополнения модели окружающей среды пользователем. Необходимо отметить, что усложнение пользовательского интерфейса произведено с целью получения максимального контроля над агентом в экспериментальных целях.

## **МОДЕЛЬ ПОЛЬЗОВАТЕЛЯ**

Модель информационных потребностей пользователя внутри агента необходима для

- обеспечения более эффективного поиска;
- обеспечения упреждающего поиска (проактивность);
- слежения за потоками информации.

Модель отображает постоянные информационные потребности пользователя (информационные запросы по слежению за источниками и запросы безотносительно источников). Она также должна содержать информацию о пользователе и об информационных источниках, которые предпочитает пользователь. Данная часть модели должна строиться и пополняться либо по желанию пользователя им самим, либо с помощью механизма отзывов (feedback) в процессе работы системы.

Важно отметить, что эффективный поиск тематической информации возможен только при наличии у агента знаний о предметной области. Мы предлагаем строить такой набор знаний в виде тематического классификатора. Классификатор содержит список тем и по каждой теме набор документов, к ней относящихся и формулировка темы в виде набора ключевых слов с весами. Пополнение такого классификатора допускается полуавтоматически на основе ответов пользователя и вручную пользователем. Хорошим вариантом для качества поиска является наличие предварительно построенного классификатора в рамках заданной узкой предметной области.

Пользовательский запрос:

- ключевые слова
- тема по классификатору
- предпочитаемый пользователем уровень документов

- тематическое сообщество, если есть
- список рекомендованных информационных источников
- директивные сроки
- тип поиска
- тип запроса (постоянный, одноразовый).

Информация о пользователе:

- постоянные пользовательские запросы (интересующие пользователя темы)
- списки рассылки (постоянно обновляемые информационные потоки) и фильтры для слежения
- история однократных пользовательских запросов
- уровень документов, предпочитаемый пользователем, уровень источников, (если надо, то по каждому из запросов)
- запрещенные источники

## **МОДЕЛЬ ОКРУЖАЮЩЕЙ СРЕДЫ**

Модель должна представлять следующие элементы окружающей среды:

- информационные источники и правила работы с ними;
- отдельные информационные ресурсы;
- средства получения информации об окружающей среде;
- средства воздействия на окружающую среду.

Пополнение модели может производиться вручную пользователем (обучение агента) с помощью специального пользовательского интерфейса и без участия пользователя, в процессе работы за счет использования так называемого расширяющего поиска.

Поиск, осуществляемый агентом можно разделить на

- поиск документов, информация о которых уже содержится в базе данных агента,
- поиск документов в рамках заданной подобласти Интернет,
- поиск документов в Интернет,
- поиск новых URL, сайтов и других информационных ресурсов.

Последний из вариантов поиска предназначен для расширения базы данных агента, тогда как предыдущие лишь могут иметь побочным эффектом выявление новых информационных ресурсов.

Получая от пользователя запрос, агент должен построить план выполнения запроса, оценить его временную стоимость и, согласовав план с пользователем, запустить его на выполнение.

Для осуществления поиска используются следующие примитивы поиска.

- Поиск в массиве известных агенту документов.
- Поиск путем использования внешних классификаторов.
- Поиск средствами одной или нескольких систем поиска по ключевым словам.
- Поиск в электронной библиотеке.
- Поиск на одном Web-сайте с встроенной системы поиска.
- Поиск документов по классификатору в рамках тематического сообщества.
- Поиск документов аналогичных заданному средствами одной или нескольких поисковых систем.

- Составление тематического сообщества по заданному URL документа.

Для выбора примитивов поиска планировщиком используется следующая информация.

- Время выполнения поиска
- Время на выдачу первого из результатов поиска
- Релевантность выдаваемой информации
- Поиск документов или поиск информационных источников
- Исходные данные для поиска
- Интерактивность, возможность автономной работы.

В дополнение к внешним примитивам поиска агент имеет встроенные средства построения поискового образа документа в виде набора ключевых слов и определения близости двух поисковых образов документов.

## **ЗАКЛЮЧЕНИЕ**

Данные предложения охватывают лишь структурную составляющую модели. В дальнейшем планируется расширение этой модели информацией о связях ресурсов по результатам наблюдения за работой пользователя.

В заключение, хотелось бы поблагодарить Костенко В.А. и Новикова А.В. за внимание и поддержку в процессе создания этой работы.

## **ЛИТЕРАТУРА**

1. Попов Э.В., Фридман Г.Р., Алгоритмические основы интеллектуальных роботов и искусственного интеллекта, М., Наука, 1976
2. C. J. Van Rijsbergen, Information Retrieval
3. Gilbert, Aparicio, The Role of Intelligent Agents in the Information Infrastructure, IBM, 1995
4. H. Stolt, Agents, filter and search engines. An evaluating survey on technologies for effective search for information from Internet resources, Umea University, 1997.
5. D. Mladenic, "Turning Yahoo into an Automatic Web-Page Classifier", Proc. of ECAI 1998.