



УДК:004.383.4.:004.822

ИЗВЛЕЧЕНИЕ ЗНАНИЙ ИЗ ИНТЕЛЛЕКТУАЛЬНЫХ ПРОГРАММ НА ЕСТЕСТВЕННОМ ЯЗЫКЕ

АБДРАУПОВА Г.Р.
izvestiya@ktu.aknet.kg

We suggest the method intended for deriving of knowledge from text on natural language, for solving the problems of optimization of streams on networking models. We consider the main stages of the working of the algorithm for deriving of information from the text. This procedure confirmed it's efficacy and capacity for work in the solving of some practical problems.

Цель исследования. Одним из основных требований к интеллектуальной программе является ее способность вести с пользователем диалог на естественном языке.

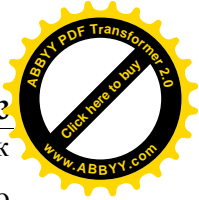
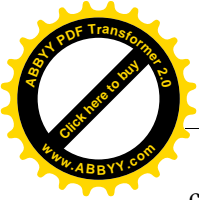
Метод исследования. “Естественность” языка общения состоит не в том, чтобы он позволял использовать весь словарь и весь арсенал синтаксиса естественного языка, а в том, чтобы вести взаимодействие с системой без какой бы то ни было предварительной подготовки пользователя.

Содержание и формы представления знаний в любом случае предполагают, что изложенное будет изучаться, анализироваться, усваиваться и использоваться для некоторых целей. Тем не менее, значительная часть материалов ориентируется на специальные нужды обучения, и это накладывает определенный отпечаток на форму и порядок изложения. В методике преподавания дисциплины учитываются особенности процессов изучения теоретических положений и выработки навыков их применения в практической деятельности. Мы можем отметить следующие моменты, связанные с преподаванием.

Процесс нахождения оптимального решения может быть представлен как последовательное сужение области поиска от всего множества исходных решений к единственному. В зависимости от специфики задачи каждый из этапов или каждое из средств могут быть конкретизированы.

Мы будем рассматривать ситуацию, при которой отсутствуют какие-либо количественные данные, которые можно было бы использовать при решении задачи, и требуется найти решение, руководствуясь только качественными оценками.

Кроме структуры самой системы знаний необходимо построить динамическую структуру процесса обучения, которая учитывает цели обучения и целевые установки обучаемых, мотивы их деятельности, а также возможности и способности.



Подобный формализм представления знаний позволяет моделировать обучение как сложную, многоаспектную систему, но при этом организовать процесс таким образом, что последовательно наращивается понимание явления обучаемым, обогащается его видение, обнаруживаются скрытые механизмы и существо происходящих изменений, развитие объекта. Для достижения этих целей важны три аспекта: организационный, классификационный и динамический.

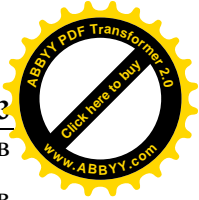
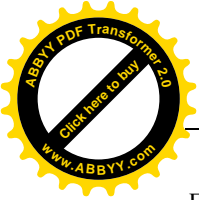
В первом, наиболее простом для понимания аспекте, изучение начинается с элементарных (неделимых на данном уровне рассмотрения и изучения) объектов. Далее, следуя отношениям «целое - часть» в базовой модели знания, изучаются более крупные образования (агрегаты). Процесс изучения продолжается до тех пор, пока не останется объектов, частями которых являются уже изученные. Отношение «целое - часть» является самым простым и наглядным. Поэтому у обучаемых не возникает барьера непонимания и им не приходится перелопачивать большие объемы материала, чтобы хоть что-нибудь понять. «Прозрение» не отодвигается на более поздние этапы обучения а наступает незаметно и постепенно, ограничиваясь пройденным материалом.

Знания, полученные в рамках организационного аспекта, служат основанием и материалом для обобщений и уяснения понятий и их родовидовых отношений в классификационном аспекте. Здесь изучение начинается «сверху», с наиболее общих понятий. Этому благоприятствуют два момента. Первый заключается в том, что уже есть достаточно обширное и в определенной мере освоенное поле знаний. Второй обусловлен спецификой определения понятий. Как правило, в определении понятия используется понятие, являющееся родовым для него. Поэтому путь последовательной конкретизации вдоль отношений «род - вид» представляется оптимальным для освоения понятий и классификационных отношений любой дисциплины.

Теперь, когда изучены структуры объектов, понятия и их отношения, может быть начато изучение процессов. Основой динамического аспекта в изучении являются причинно-следственные связи. Они позволяют показать и объяснить изменения и развитие, которое мы наблюдаем в действительности и, может быть, которым хотели бы управлять. Динамический аспект позволяет нам сделать шаг и в этом направлении. Не только понимать и осознавать, но и эффективно действовать человек может только на основе целостного знания.

Язык общения должен обеспечивать пользователя простыми средствами однозначной формулировки задачи в рамках конкретных приложений. Однако научить ЭВМ понимать естественный язык очень трудно, так как предложения в человеческой речи часто бывают неоднозначными и скорее подразумевают какие-то факты, чем непосредственно констатируют их [1].

Проблема приобретения знаний компьютерными системами является ключевой как с позиций понимания устройства человеческого знания, так и с позиций производства новых компьютерных систем.



Одним из источников знаний для интеллектуальных систем является множество текстов предметной области. Из всего этого множества можно выделить класс так называемых “текстов исходных данных”. Анализ этого класса текстов показал, что они состоят из набора более или менее однотипных конструкций. Еще одной отличительной чертой выделенного класса текстов является то, что информация, заложенная в них, носит декларативный характер и не подлежит занесению в базу знаний [2]. Тем не менее иметь возможность извлечь эту информацию необходимо, так как не всегда возможно или удобно задать исходные данные решающей программе в понятном ей виде.

Повторение в тексте одних и тех же конструкций делает возможным научить ЭВМ автоматически извлекать нужную информацию из каждой конструкции. Тщательно проанализировав все конструкции языка, используемые в данном типе текста, можно сделать вывод о том, что все конструкции можно разбить на группы, количество которых конечно и невелико. А, следовательно, для каждой группы конструкций можно разработать свою стратегию поведения программы. При правильном разбиении на группы различия между конструкциями внутри группы незначительны и существенно не влияют на поведение программы.

Распознать каждую группу можно по некоторому слову, наличие которого в данной конструкции обуславливает расположение остальных элементов. Рассмотрим все вышесказанное применительно к решению задачи оптимизации потоков на сетевых моделях.

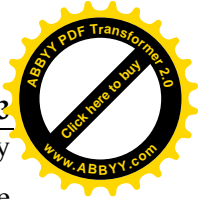
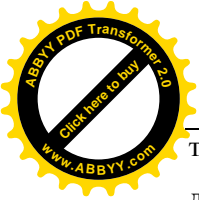
При решении задач этого класса модель представляется в виде графа, который задается в виде матрицы весовых коэффициентов. Для графов с небольшим (порядка 10) количеством вершин не составляет особых сложностей ввести эту матрицу вручную, но для большего числа вершин становится легче описать этот граф текстом. Это происходит потому, что при описании графа текстом необходимо указать только существующие связи, а не все возможные (т.е. и несуществующие тоже).

Информация, которую мы будем искать в этом тексте - это длина дуги между двумя вершинами.

Рассмотрим пример текста, описывающего граф:

“ Граф включает 5 вершин. Первая и вторая вершины соединены с третьей, которая связана, в свою очередь, с четвертой и пятой вершинами. Пятая вершина связана дугой с первой вершиной. Длины дуг следующие: между первой и третьей вершиной - 5; между второй и третьей вершиной - 6; между третьей и четвертой вершиной - 3; между третьей и пятой вершиной - 8. Длина дуги от пятой вершины до первой равно 10. ”

В этом тексте представлены некоторые варианты конструкций, а также видны некоторые проблемы, с которыми сталкивается разработчик программы. Основные из них: неполная информация в одной конструкции (“Пятая вершина связана дугой с первой вершиной”) и сложноподчиненное предложение, составными частями которого являются конструкции с неполной информацией (“Первая и вторая вершины связаны с третьей, которая связана, в свою очередь, с четвертой и пятой вершинами”). Первая проблема возникает из-за



того, что программа может одновременно рассматривать только одно предложение. Поэтому для решения этой проблемы необходимо анализировать весь текст до конца. И только после этого проводить проверку на наличие данных и, в случае отсутствия каких-то данных, обращаться за помощью к пользователю. Для решения второй проблемы необходимо обеспечить возможность разбиения такого предложения на два простых. Дальнейшие действия совпадают с решением первой проблемы. Но возникает необходимость научить программу различать сложноподчиненные предложения, так как могут встретиться предложения, не требующие разбиения (например, “вершины i и j соединены между собой дугой, длина которой равна L ”). Конечно, было бы гораздо проще и удобнее ввести жесткий формат текста, т.е. создать текст, состоящий только из простых предложений с полной информацией в них. Но тогда пришлось бы подготавливать каждого пользователя, а это противоречит основному принципу, высказанному в начале статьи: “Естественность языка ... состоит ... в том, чтобы он позволял вести взаимодействие с системой без какой бы то ни было подготовки пользователя”.

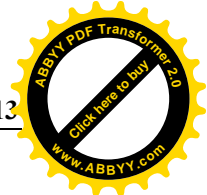
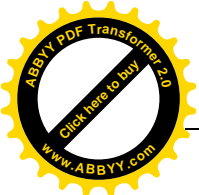
На основании всего вышесказанного можно предложить следующий алгоритм выделенного класса текстов:

- За всеми порядковыми числительными, где это необходимо, вставить относящиеся к ним существительные в соответствующем падеже.
- Заменить все числительные, написанные словами, на числительные, написанные цифрами.
- Заменить необходимые сложноподчиненные предложения простыми.
- Взять первое предложение текста.
- Искать в нем первую конструкцию.
- Если конструкция обнаружена, то извлечь информацию и удалить конструкцию из предложения.
- Если обработаны не все конструкции, то взять следующую и перейти на шаг 6.
- Если обработан не весь текст, то взять следующее предложение и перейти на шаг 5, иначе конец.

Первые три пункта алгоритма предназначены для приведения текста к некоторой стандартной форме для облегчения распознавания.

Предложенный алгоритм реализован на искусственном языке. Применение алгоритма при решении некоторых практических задач обработки текстов на естественном языке подтвердило его работоспособность и высокую эффективность.

В заключение следует признать, что предложенный метод не лишен недостатков. Вместе с тем к несомненным достоинствам предложенного метода относится скорость, с которой обрабатывается текст.



1. Г. С. Осипов. Приобретение знаний интеллектуальными системами. - М.: Наука. 1997.
2. Статические и динамические экспертные системы./ Э. В. Попов, И. Б. Фоминых, Е. Б. Кисель, М. Д. Шапот. - М.: Финансы и статистика. 1996.