ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ С ЕСТЕСТВЕННО-ЯЗЫКОВЫМ ИНТЕРФЕЙСОМ

Сироткин А.В.
Кафедра информационных технологий автоматизированных систем
Научный руководитель: Герман О.В., кандидат технических наук, доцент
email: salexvlad@gmail.com

***Аннотация* – Рассматриваются современные подходы к обработке текстов на естественном языке.**

Ключевые слова: ИС, анализ, информация

Использование интеллектуальной системы с естественно языковым интерфейсом (ИС) должно улучшить поиск. Другими словами, в идеале, ИС должна сама формировать ответ из своей базы знаний (БЗ). А как формировать БЗ? Для начала, в неё должны быть занесены ответы на часто встречаемые вопросы. А как хранить всё остальное? Для начала вводимая информация будет подвержена нескольким уровням анализа:

1. Фонетический – на этом этапе производится преобразование букв слова в фонемы. Тогда слова «приготовиться» и «пригатовиться» будут равнозначны. На этом этапе отсекается ряд ошибок.
2. Морфологический и синтаксический - выполняется анализ слова и совокупности слов. Выполняются различные исправления ошибок. Определяются части речи и члены предложения, чтобы затем построить структуру предложения. Принимаются различные допущения, в случае неконкретных предложений (неполных, незаконченных).
3. Семантический – словам, словосочетаниям, может даже предложениям ставится в соответствие какая-то константа. И на основе констант, строится модель полученной информации.
4. Модельный – на этом этапе, из собственных данных БЗ и полученной модели строится новая обобщенная модель. И тогда с помощью этой модели, система может делать какие-то заключения и выдавать ответ.

Фонетический, морфологический и синтаксический анализ достаточно хорошо проработаны. Основные проблемы связаны с разбором неполных предложений, что вызывает трудности с семантическим анализом, когда нельзя точно определить смысл предложения. К примеру, «Федя играет со своей женой Леной два раза в неделю. Петя – тоже». Однозначно смысл предложения определить нельзя. Или предложение из анекдота «Из окна дуло», в этом утверждении определить часть речи слова «дуло» нельзя. Такие неоднозначности разрешаются только с раскрытием контекста. Для семантического анализа должна быть разработана система соответствий, своего рода переводчик-словарь, которая по определенным правилам ставит в соответствие слову, предложению определенную константу.

Реализовать первый уровень анализа позволяют большинство языков программирования. На входе последовательность букв слова, а на выходе последовательность его фонем. Для реализации второго уровня анализа разработаны следующие решения: mystem (Yandex), TreeTagger, также ntlk (Python). Алгоритм для морфологического анализа может быть следующим: формируются начальные данные по схеме слово, составляющие морфемы – часть речи, далее определяются начальные и последние несколько букв (фонем), части речи предыдущих слов предложения. Затем используем алгоритм обучения, к примеру, метод опорных векторов (SVM), который принимает данные и выдает классификацию по заранее заданным правилам. В итоге, SVM построит модель, на базе большого количества информации, которая в большинстве случаев корректно определяет часть речи. Для синтаксического анализа целесообразно использование генератора парсеров., а именно GLR парсеров. Для реализации семантического анализа можно использовать искусственные логические языки, например lojban. Плюсы языка - свобода выражения, точность речи и однозначность каждого предложения. И на заключительном этапе данные представлять в виде правил и фактов, используя программную среду для разработки экспертных систем CLIPS. В качестве альтернативы можно использовать ОО язык INFORM. Основные сложности на заключительном этапе – это процесс интеграции моделей., разрешение противоречивости.

Тогда запрос пользователя проходит выше перечисленные уровни анализа, после чего в базе составляется модель (ответ), которая переводится в естественный язык. Ответ будет то

Сферы применения ИС, кроме поисковых различны: от чат-ботов до голосовых интерфейсов.

[1] U.Eco The role of the reader – Moscow (Russia), 2005
[2] www. habrahabr.ru/
[3] www. wikipedia.org