

оценивания и сравнения вариантов, и осуществляется выбор конкретного варианта проектируемой распределенной ИУС.

На третьей фазе реализуется выбранный проект системы. Решаются задачи оптимизации, субоптимизации (объяснение того, почему наилучшее решение не может быть получено). Оценивается сложность задач, и при необходимости проводится упрощение реальности, разрешаются конфликты. Затем происходит оценка результатов, полученных от внедренного проекта распределенной ИУС, а также определяют необходимые доработки процесса проектирования распределенной ИУС.

Последовательность фаз можно рассматривать как унифицированный метод проектирования распределенных РИУС. Каждая из фаз подлжит алгоритмизации, что составляет общий процесс проектирования РИУС и полностью отражает процесс структуризации работ при проектировании.

#### БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Самсонов В.С.* Автоматизированные системы управления / Учеб. Для учащихся энерг. спец. техн. – М.: Высшая школа, 1991.
2. *Энкарначо Ж., Шлехтендаль Э.* Автоматизированное проектирование. Основные понятия и архитектура систем / Пер. с англ. – М.: Радио и связь, 1986.
3. *Недужко А.М., Крючек М.И.* Методология современного проекта // Материалы Международной научной конференции «Проектирование новой реальности» (ПНР-2007). Ч. 1. – Таганрог: Изд-во ТТИ ЮФУ, 2007.
4. *Косенко Е.Ю., Макаров С.С., Финаев В.И.* Методы моделирования и проектирования распределенных информационно-управляющих систем. – Ростов н/Д.: Изд-во Рост. ун-та, 2004. – 199 с.
5. *Поспелов Г. С. Ириков В.А.* Программно-целевое планирование и управление. – М.: Радио и связь, 1976.
6. *Тимченко А.А.* Эффективность проектных процессов и качество проектных решений. – Киев: Общество «Знание», 1982.
7. *Шеверов Д.Н.* О методических основах автоматизации проектирования технических систем//Автоматизация проектирования. – М.: Машиностроение, 1986 Вып. 1. – С. 188-202.
8. *Маклаков С.В.* Создание информационных систем с AllFussion Modeling Suite. – М.: Диалог-МИФИ, 2003. – 432 с.

УДК 621.396

**Р.Ю. Вишняков**

#### **СЕМАНТИКА И СМЫСЛОВЫРАЗИТЕЛЬНОСТЬ В ИНФОРМАЦИОННОМ ПОИСКЕ**

В связи с лавинообразным ростом объемов информации все более обостряется актуальность разработки эффективных поисковых средств. Часто вопрос состоит уже не в том, чтобы найти некоторый источник данных с интересующей пользователя информацией, а в том, чтобы отобрать из множества источников наиболее отвечающие поисковому запросу документы. На сегодня сложилось множество подходов в решении данной проблемы, что способствовало формированию и становлению области знаний, называемой информационным поиском.

В широком смысле слова в поисковую проблематику включаются непосредственно вопросы поиска документов, обработки результатов, моделирования, классификации, кластеризации, фильтрации, проектирования архитектур поисковых систем, пользовательских интерфейсов, поисковых языков пр. Несмотря на многообразие задач, все они касаются, так или иначе, интерпретации и моделирования семантики языковыми средствами.

Особенностям языка, как средству отражения объективной картины мира в контексте философского соотношения «конкретное-абстрактное», посвящено много исследований, начиная от древнего мира (Платон. Кратил. – Соч.: В 3-х т., М.: Мысль, 1968, т.1), заканчивая современными исследованиями, например, [1]. Более того, сама логика человеческого познания построена на движении от конкретного к абстрактному путем выявления наиболее общих закономерностей в явлениях окружающего мира. Метафорически человеческое «Древо знаний» отражает классифицирование явлений окружающего мира в восходящий ряд обобщающих понятий. [2]. Так, можно говорить о классификациях понятий развивающихся отраслей знаний (биологии, химии, математики, информатики и т.д.). Не вдаваясь в саму структуру таких классификаций, отметим только, что они отражают различные отношения между сущностями вещей (понятиями), находящимися вне человеческого сознания. Такие отношения явлений внешнего по отношению к человеку мира называются *парадигматическими* и они не зависят от выражающего их языка. Например: родовое отношение *отец-сын*, видовое отношение *дерево-вишня* и т.д., на каких бы языках они не выражались, не перестают быть таковыми.

В тоже время, если бы человеческий язык был устроен так, что каждому новому смыслу давал бы новое название, он разрастался бы неимоверно. Однако словарный запас языка (лексический состав) представляет собой ограниченное множество слов, которые с точки зрения выражаемого смысла являются многозначными (полисемичными). В тоже время, слова, связываясь в цепочечные структуры (предложения, абзацы, разделы, книги и пр.), позволяют задавать новый смысл путем вычленения из множества значений слов их более узких смысловых значений (сужение полисемичности) и их синтетического объединения. Например, в предложениях *«отчий дом, что может быть милее и роднее»* и *«каменный дом, как крепость, монументальная и неприступная»* речь идет о разных смысловых значениях цепочечных структур, а само слово *дом* участвует в них в разных смысловых значениях.

Очевидно, что только человеческое сознание, интерпретируя такие цепочечные структуры, выявляет смысл. Таким образом, смысл (семантика) является свойством человеческого сознания, более того, окрашена конкретной личностью.

Пусть  $L$  – некоторый язык и в нем существует последовательность  $\alpha$  слов  $x_1, x_2, \dots, x_n$ , представляющая собой некоторое осмысленное выражение (словосочетание, предложение, абзац, статью, книгу и т.д.), т.е.  $\alpha = x_1, x_2, \dots, x_n$ . Пусть  $S(x_i)$  – множество смыслов (семантических значений) слова  $x_i$ , тогда смысл (семантическое значение) выражения  $\alpha$  можно определить следующим функционалом  $S(\alpha) = \Phi(x_1, x_2, \dots, x_n)$ . Данный функционал вычленяет из множества смыслов слов определенные семантические значения и на множестве таких значений строит значение смысла всего выражения  $\alpha$ . Характерной особенностью  $S(\alpha)$  является то, что он представляет новый смысл в человеческом сознании, в общем случае не разлагаемый на смыслы отдельных слов и не сводящийся ни к

одному из них. Скорее всего, совокупность узких смыслов слов позволяет высветить (проявить) в человеческом сознании новый смысл  $S(\alpha)$ .

В общем случае, о виде данного функционала, о его работе в каждой конкретной ситуации практически ничего не известно. Но, в тоже время, человеческое сознание при восприятии предложений языка достаточно точно его отрабатывает и определяет. Говоря о физической интерпретации природы понимания смыслов [2], автор предполагает, что выражение – это некий своеобразный фильтр, который из универсума смыслов выделяет заданный.

Обратимся теперь к отдельным словам. Пусть задано слово  $\beta$  и множество его смысловых значений  $S(\beta) = \{s_1, s_2, \dots, s_n\}$ . Графически это можно представить в виде множества или графика, изображенных на рис.1, а и рис.1, б соответственно. В случае (рис. 1, а) каждая точка в круге представляет отдельное смысловое значение слова. В случае (рис. 1, б) точки оси абсцисс  $S(\beta)$  представляют множество смыслов слова  $\beta$ , которые, для наглядности восприятия, упорядочены по частоте их употребления  $f(S_i)$  относительно наиболее употребляемого смыслового значения. Очевидно, что данный график в некотором смысле является субъективным, но, в то же время, он отражает реальную картину.

Обращаясь к языку как объекту, отметим, что он сам представляет собой множество предложений. В предложениях языка между словами существуют определенные отношения, их называют синтагматическими. Формально язык представляется алфавитом, словарным составом (лексикой) и множеством правил (грамматикой) катенации слов в цепочечные (линейные) структуры, которые являются средствами метаописания языка. Обычно грамматику делят на морфологию и синтаксис. Морфология задает способы построения и изменения слов, а синтаксис – соединения слов в предложения. Синтагматика в естественных языках, например, типа русского языка, задается в описательном виде, в формальных языках, какими являются языки программирования, используются, так называемые, порождающие грамматики.

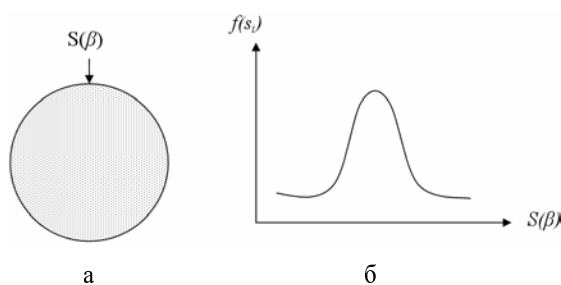


Рис. 1. Задание точки: а) множеством; б) графиком

Отметим, что с формальных позиций синтагматика языка никаким образом не связана со смысловыразительностью, поэтому грамматически правильно построенные слова и предложения могут быть абсолютно бессмысленными. Но в тоже время именно эта особенность языка делает его открытым к развитию и расширению.

Рассмотрим, как смысловыразительные особенности языка отражаются в поисковой проблематике.

Введем понятие морфологического подобия. Пусть задан язык  $L$ . Под морфологическим подобием для слов  $\alpha$  и  $\beta$  будем называть характеристику  $p(\alpha, \beta)$ , которая принимает значения на интервале  $[0, 1]$ . В случае, когда  $p(\alpha, \beta) = 1$ , слова равны с точностью до буквы. В случае, когда  $0 < p(\alpha, \beta) < 1$ , слова подобны по каким-либо признакам. Например, имеют одну и ту же основу, или являются однокоренными, или и т.д. имеют какие-либо общие морфологические признаки. В случае, когда  $p(\alpha, \beta) = 0$ , общие морфологические признаки у слов отсутствуют. Обсуждение формирования характеристики морфологического подобия выходит за рамки данной работы, поэтому в нашем случае будем считать, что эта характеристика формируется внешним экспертом.

Для дальнейших рассуждений примем следующие обозначения. Пусть  $\beta$  – поисковый запрос, а  $D = \{d_1, d_2, \dots, d_n\}$  – множество документов или поисковое документное пространство, по которому осуществляется информационный поиск.

Информационный поиск на основе морфологического подобия сегодня является наиболее распространенным, в особенности, это относится к Интернету. Поисковые запросы здесь составляются в виде отдельных слов или их логических комбинаций, а сама поисковая процедура состоит в отыскании в текстах документов вхождений слов поискового запроса на условиях морфологического подобия. Очевидно, что задавая характеристическую функцию морфологического подобия, можно реализовать различные стратегии и виды поиска. Например, сюда включаются все виды полнотекстового поиска.

Однако характерная особенность поисков на основе морфологического подобия состоит в том, что семантическая составляющая запроса здесь не вычленяется и не учитывается, она предполагается равной множеству смысловых значений входящих в поисковый запрос слов. Например, если поисковый запрос образует слово, то неявно предполагается, что ищутся все документы по всем смысловым значениям. Если в запрос входят слова  $x, y, z$ , то по всем документам ищется смысловое значение  $S(x) \cup S(y) \cup S(z)$ . Попытка понизить шумовую составляющую в результатах поиска приводит к введению внешних ограничений на выдачу, которые как бы моделируют семантику запроса. Так, выдвигается следующая гипотеза о релевантности документа запросу. Из двух документов более релевантен запросу тот, у которого слово запроса встречается большее число раз. К сожалению, уязвимость такой гипотезы более чем очевидна, что мы и наблюдаем в Интернет.

Использование принципа координации [3] наоборот предполагает использование только семантической составляющей понятий. Морфологическая составляющая здесь играет вспомогательную роль, она используется для идентификации понятий и фактически является его именем (идентификатором). Координация предполагает, что предварительно все понятия проклассифицированы и организованы в упорядоченные иерархии по каким-либо признакам. В соответствии с этими классификациями для всех документов, введенных в информационную систему, построены поисковые образы, отражающие их семантические содержания, или, иначе говоря, построен семантический тезаурус.

Лексика языка поисковых запросов представляет собой только множество имен понятий, а предложение поискового запроса может образовывать одно или несколько таких имен. Если в поисковый запрос входят слова  $x, y, z$ , то он имеет четкое смысловое значение  $S(x) \cup S(y) \cup S(z)$  и все документы, соответствующие данному семантическому значению, попадут в выдачу.

Как правило, информационные системы, основанные на координатных принципах, являются узкоспециализированными, поскольку расширение поискового языка новым словом всегда требует включения связанного с ним понятия в систему классификаций, модификации семантического тезауруса и пополнения документной базы. В координатных языках большого лексического объема усиление смысловыразительности достигается за счет введения характеристик роли и связи, однако это не решает в принципе проблему языковой синтагматической бедности.

Таким образом, анализ языка с позиции знак-смысл, введение характеристики морфологического подобия позволило оценить языковые средства, используемые в информационном поиске, и построить шкалу их смысловыразительности. Крайними точками на этой шкале являются поисковые средства морфологическом подобия и координации. Очевидно, что в каждом из них используется только одна из составляющих пары знак-смысл, в то время, как сама эта пара обладает внутренним единством, выраженном в функционале  $S(\alpha)$ , который, в свою очередь, учитывает синтагматику языка.

Использование методов и средств искусственного интеллекта, включающих базу знаний и логический вывод, моделирующих человеческое рассуждение, можно рассматривать как поиск промежуточных решений на этой шкале в развитие координатного принципа. В тоже время, усложнение поискового языка и введение в него элементов синтаксического и семантического анализов также можно рассматривать поиск решений на это шкале в направлении от морфологического подобия.

#### БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Налимов В.В.* Вероятностная модель языка. О соотношении естественных и искусственных языков. – М.: Наука, 1979. – 303 с.
2. *Яновская С.А.* Проблемы введения и исключения абстракций более высоких порядков (чем первый) // В кн.: Яновская С.А. Методологические проблемы жизни. – М.: Мысль, 1972.
3. *Черный А.И.* Введение в теорию информационного поиска / Монография. – М.: Наука, 1975. – 239 с.