

УДК 004.5

**ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ МНОГОЗНАЧНОЙ ПОДСКАЗКИ
В ИНТЕРФЕЙСЕ ПОЛЬЗОВАТЕЛЯ СИСТЕМЫ ПОИСКА
ПО КЛЮЧЕВОМУ СЛОВУ**

В.А. Литвинов, С.Я. Майстренко, И.Н. Оксанич

Институт проблем математических машин и систем НАН Украины

e-mail: maistrsv@ukr.net

Как показано в теоретических моделях [1, 2] уменьшение значения m объема порции пошаговой подсказки в системе поиска по ключевому слову (КС) ведет к уменьшению среднего количества просматриваемых слов $\bar{M}$ и снижению трудоемкости N ввода КС.

Минимальные значения $\bar{M}$ и $\bar{N}$ достигаются при $m=1$ и близки к соответствующим значениям прицельной подсказки [3]. Однако, при этом уменьшаются до минимума возможности замещения ключевого слова неким потенциальным синонимом (ПС), который может появиться в порциях подсказки раньше, чем будет введен детерминант ключевого слова – минимальное количество начальных символов, однозначно его определяющих. В дополнение к более раннему появлению, ПС может и точнее передавать смысл запроса пользователя. Иными словами, в пошаговой подсказке заложены механизмы как подсказки набора символов КС, так и подсказки уточненного смысла КС. Для простых информационно-поисковых систем (ИПС) с близким к однозначному соответствию «ключевое слово – ресурс» подсказка уточненного смысла если и возможна, то не имеет особого практического значения. Для ИПС, где это отношение имеет более сложный, многозначный характер (типа «многие к многим») ситуация оказывается иной, и возможность замещения КС синонимом должна давать ожидаемый двоякий эффект.

Цель эксперимента состояла в натурном моделировании процесса подсказки в реальной ИПС с многозначными отношениями «ключевое слово – ресурс» для установления ориентировочных зависимостей и количественных соотношений между существенными (в контексте рассматриваемого вопроса) исходными и результатными параметрами пошаговой подсказки.

В глобальной ИПС GOOGLE (браузер IE-8) осуществлялся ввод 32-х произвольно выбранных из [4] КС, состоящих из одного слова длиной n равной 7-10 символов и 32-х двукратных словосочетаний, разделенных пробелом (2-словосочетаний), суммарной длиной 15-20 символов.

Общения с GOOGLE проходило в двух режимах:

А – использование синонимического замещения.

Если в порциях подсказки появлялся ПС, он выбирался в качестве искомого ключевого слова. Если среди 10 первых ресурсов, предоставленных по запросу не находилось удовлетворительного (в смысле приемлемой релевантности), ввод КС повторялся с игнорированием ПС.

В – игнорирование синонимического замещения.

Процесс ввода символов и анализа порций подсказки продолжался до появления в подсказке именно конкретного КС.

В качестве критериев оценки эффективности режимов кроме реальных значений $\bar{v}$ (среднего количества введенных символов), $\bar{M}$ и индикативных значений N дополнительно фиксировался порядковый номер $k < 10$ ресурса предоставляемого системой и удовлетворяющего сделанный запрос.

Общие результаты экспериментов следующие.

1. Одиночные слова. Из 32-х слов потенциальные синонимы нашлись для 12. Из них удовлетворительные результаты получены только для 5 замещений: <архивация>→<архиватор>, <бизнес-модель>→<бизнес-моделирование>, <браузер>→<браузеры>, <веб-квест>→<веб-квесты>, <драйвер>→<драйвера>.

Примеры неудачных замещений: <загрузчик>→<загрузки>, <запрос>→<запросы google>, <канал (ТВ, радио)>→<канал футбол>.

2-словосочетание. Из 32-х словосочетаний нашлись и были проверены 17 ПС; удовлетворительные результаты получены для 10 замещений: <векторное представление>→<векторная графика>, <виртуальное соединение>→<виртуальное сетевое подключение>, <динамический объект>→<динамический массив>, <императивный язык>→<императивное программирование>.

Остальные удачные замещения – это подстановка множественного числа вместо единственного, или наоборот: <ключевое слово>→<ключевые слова> и т.п.

Примеры неудачных замещений: <витрина данных>→<витрина>, <искусственный интеллект>→<искусственный разум>.

Обобщенные количественные характеристики результатов проведенного натурного моделирования сведены в таблице.

Таблица. Обобщенные результаты натурного моделирования

Ключ. слово	$\bar{n}$	Режим	m						$\frac{H(10)}{H(4)}$	$\bar{k}$
			10			4				
			$\bar{v}$	$\overline{M}$	H	$\bar{v}$	$\overline{M}$	H		
1.	8.7	A	5.22	44.87	21.24	5.53	21.81	12.35	1.72	3.09
		B	4.84	41.84	19.89	5.16	18.53	10.9	1.82	1.69
2.	18.3	A	8.91	83.53	54.25	10.66	41.00	30.14	1.80	3.28
		B	9.25	87.69	56.81	10.65	40.94	30.10	1.88	1.16

Данные таблицы показывают, что в проведенном эксперименте использование потенциальных синонимов не привело к ожидаемому сокращению трудоемкости ввода КС: значения $\bar{M}$ и N режима А по сравнению с В оказались примерно равны для словосочетаний и даже выше для одиночных слов. Более того, для режима А оказались заметно выше и значения $\bar{k}$, характеризующие «вторичную» трудоемкость поиска релевантного ресурса из числа представленных поисковой системой.

Приведенные данные в силу разных причин не могут, конечно, служить исчерпывающим основанием предпочтительности малых значений m ($3 \div 4$) при многозначном соответствии «КС-ресурс», однако они говорят в пользу такого выбора. Для систем поиска с однозначным соответствием КС (например, идентифицирующие коды) значение m может быть еще меньше – вплоть до 1 (или использования «прицельной» подсказки [3]).

Литература

1. Логико-вероятностная модель пошаговой подсказки в интерфейсе пользователя поисковой системы по ключевому слову / Г.Е. Кузьменко, В.А. Литвинов, С.Я. Майстренко, И.Н. Оксанич // Математичні машини і системи. - 2011. - №2. - С. 41-49.
2. О выборе объема порции подсказки в задаче поиска по ключевому слову / В.А. Литвинов, С.Я. Майстренко, И.Н. Оксанич // Матеріали науково-практичної конференції з міжнародною участю «Системи підтримки прийняття рішень. Теорія і практика». – Київ. – 2011. – 6 червня. - С.80-81.
3. Интеллектуализованный интерфейс пользователя информационно-поисковой системы в задаче поиска по ключевому слову («образцу») с упреждающей подсказкой / Г.Е. Кузьменко, В.А. Литвинов, С.Я. Майстренко, И.Н. Оксанич // Математичні машини і системи. - 2011. - №1. - С. 61-71.
4. Толковый словарь по информатике.
<http://www.bitoman.ru/download/slow/112384/9d4c2.html>.