

УДК 681.3

МОДЕЛЬ ОЦЕНКИ ПАРАМЕТРОВ ИНТЕРФЕЙСА ПОЛЬЗОВАТЕЛЯ В ЗАДАЧЕ ПОИСКА ПО КЛЮЧЕВОМУ СЛОВУ С УПРЕЖДАЮЩЕЙ ПОДСКАЗКОЙ

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

Институт проблем математических машин и систем НАН Украины

e-mail: maistrsv@ukr.net

1. Типовая задача, реализующая доступ к информационным ресурсам СППР заключается в задании некоего образца (ключевого слова) и его нахождения в базовом словаре (БС) слов-эталонов.

Упреждающая подсказка состоит в том, что по мере ввода начальных символов образца, - вплоть до ввода «детерминанта» [1], однозначно определяющего образец, - прогнозируются возможные варианты искомого эталона и предоставляются пользователю (Рис.1).

Область приложения таких задач весьма широка – от простых систем учета товаров на складе до поисковых систем мультимедийных Web-ресурсов СППР.

2. Общий алгоритм интерфейсов по схеме рис.1 заключается в следующем.

Пользователь последовательно вводит символы a_1, a_2, \dots, a_n образца (начиная с первого, старшего). На каждом шаге из БС в текущий справочник (возможно, виртуальный) помещаются слова с одинаковыми значениями символов $a_1, a_1a_2, \dots, a_1a_2 \dots a_i$.

На схеме рис.1 показана ситуация для $i = 2$. Из свойства лексикографической упорядоченности слов-эталонов ясно, что $m_{i-1} \rangle m_i$, т.е. область поиска образца сужается по мере ввода символов a_1, a_2, \dots до тех пор, пока сохраняется неравенство $m_i \rangle m$, где m - некоторое заданное значение объема визуально обозримой страницы текущего справочника (при вводе детерминанта $m = 1$).

3. Для оценки производительности интерфейса [2,3] определяющую роль играет значение среднего количества символов \bar{v} , вводимых до достижения $m_i \leq m$. Обозначим такое событие, определяющее окончание процесса ввода символов, через $S(i)$.

Общее выражение для \bar{v} может быть записано в следующем виде:

$$\bar{v} = \sum_{i=1}^n i \cdot p(i), \tag{1}$$

где $p(i)$ - безусловная вероятность события $S(i)$.

В свою очередь

$$p(i) = p_i \prod_{k=1}^{i-1} (1 - p_k), \tag{2}$$

где $p_{i,k}$ - условная вероятность события $S(i)$ на шаге i при условии, что оно не произошло на предыдущих шагах $(i-1), (i-2), \dots, 1$.

Предполагая случайным распределение N реально существующих слов БС среди q^n всевозможных значений комбинаций n символов в алфавите q , рассмотрим следующую модель.

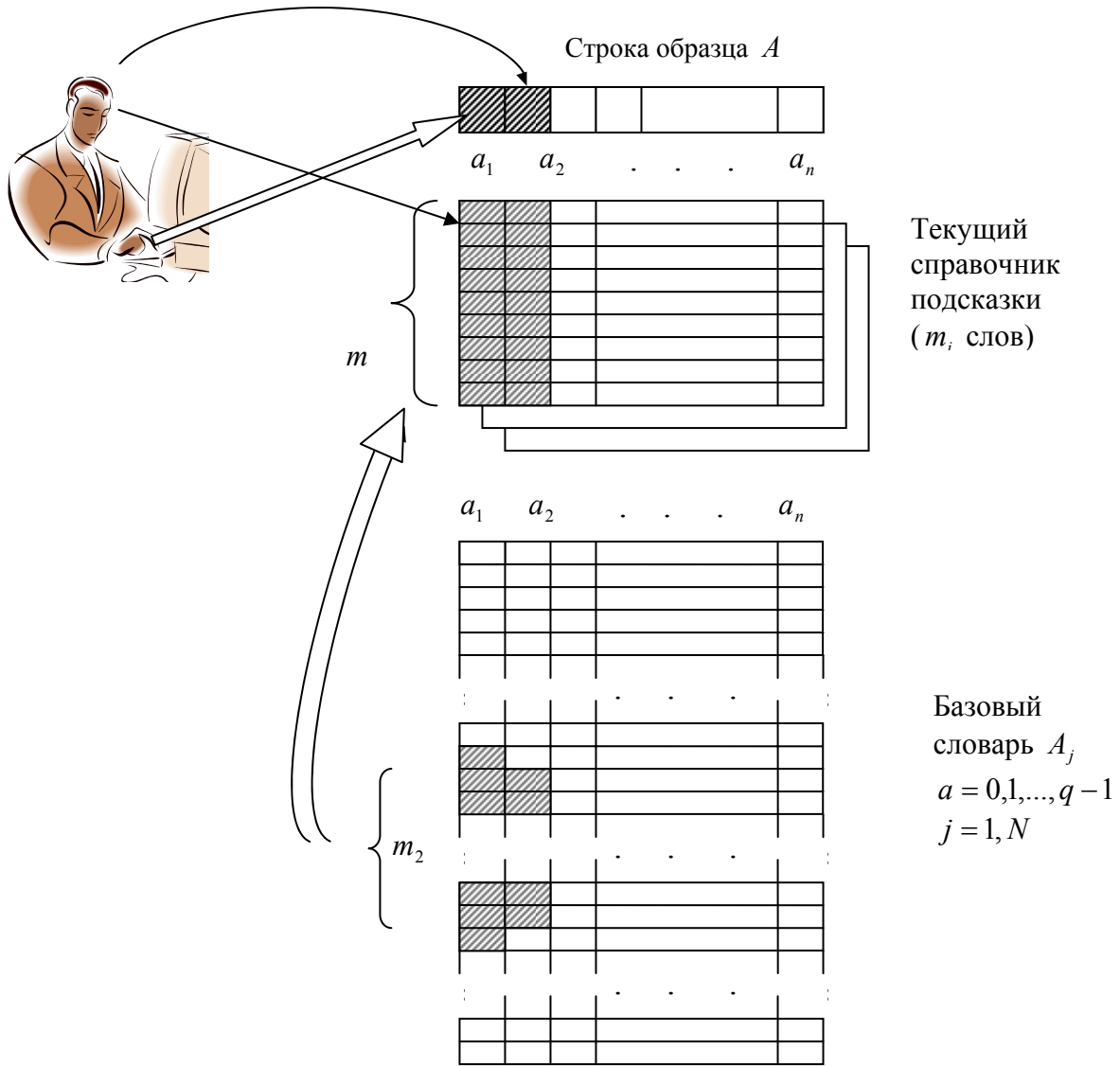


Рис. 1. Общая схема интерфейса с упреждающей подсказкой

В некую урну "с узким отверстием" вбрасывается M шаров. Вероятность того, что очередной шар попадает в отверстие урны равна $r = N/q^n$. Поскольку эта вероятность не зависит от результатов предыдущих бросаний, общие вероятностные результаты вбрасываний могут быть описаны схемой и соотношениями процесса независимых испытаний Бернулли, в соответствии с которыми вероятность $P[\overline{(o, x)}, r, M]$ получить в точности x или менее удачных исходов равна

$$P[\overline{(o, x)}, r, M] = \sum_{g=0}^x C_M^g \cdot r^g \cdot (1-r)^{M-g}, \quad (3)$$

В соответствии с такой интерпретацией для шага i положим $M = q^{n-i}$, а при определении значений p_i учтем следующее.

В случае потенциально положительного результата поиска в БС есть искомый образец, и среди m_i слов одно имеет не случайное, а детерминированное происхождение. Это означает, что один шар кладется в урну заранее, а $0, 1, \dots, m-1$ шаров попадают случайным образом, определяя удачные исходы испытаний Бернулли (т.е. в выражении (3) следует принять $x = m - 1$).

Таким образом,

$$p_i = P\left[\overline{(0, m-1)}, r, q^{n-i}\right], \quad (4)$$

Полагая $\prod_{k=1}^n (1 - p_i) = 0$ (т.е. процесс определенно заканчивается при $i \leq n$): запишем следующее расчетное выражение для искомого значения \bar{v}

$$\bar{v} = \sum_{i=1}^n i \cdot P\left[\overline{(0, m-1)}, r, q^{n-i}\right] \prod_{k=1}^{i-1} \{1 - P\left[\overline{(0, m-1)}, r, q^{n-i}\right]\} \quad (5)$$

4. Для проверки адекватности модели (1) – (5) применительно к реальным распределениям значений слов-эталонов в БС и оценки времени обработки БС построена имитационная модель (ИМ) процесса ввода символов образца и анализа текущих значений v .

Блок-схема ИМ приведена на рис.2

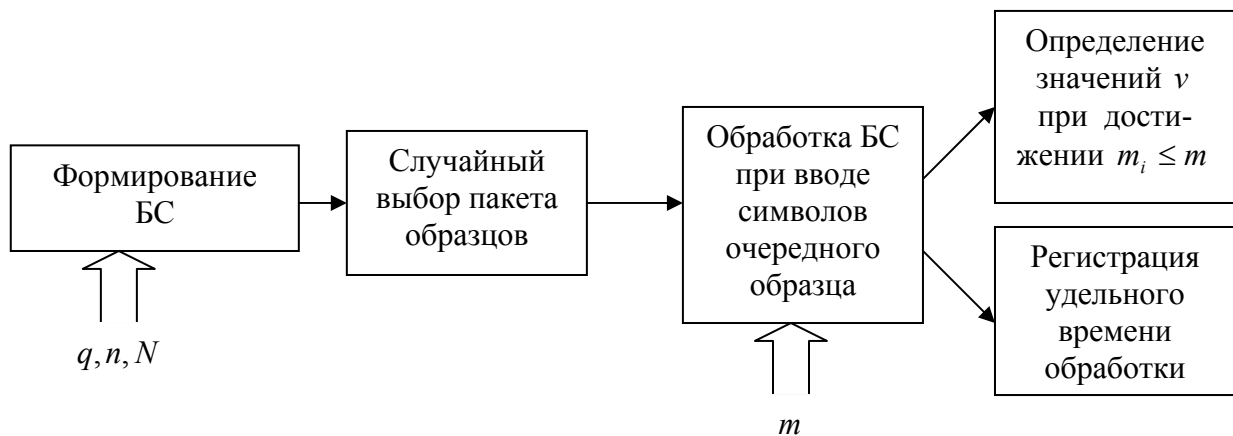


Рис.2. Блок-схема ИМ

ИМ запрограммирована в среде разработки Delphi, моделирование проведено на компьютере P4(2,8 ГГц).

Для ускорения обработки БС предложен алгоритм, основанный на идеях итерационного поиска.

В таблице приведены расчетные значения $\bar{v}_{\text{расч}}$, экспериментальные значения $\bar{v}_{\text{эксп}}$, удельные значения времени обработки одного образца τ для принятых значений исходных параметров.

Таблица. Расчетные и экспериментальные данные

m	r	$q = 32, n = 8$			$q = 10, n = 12$		
		$\bar{v}_{\text{расч}}$	$\bar{v}_{\text{экспк}}$	τ	$\bar{v}_{\text{расч}}$	$\bar{v}_{\text{экспк}}$	τ
10	10^{-6}	3,99	4,02	11,0	5,54	5,55	38,6
	10^{-8}	2,63	2,52	<1,0	3,54	3,55	<1,0
1	10^{-6}	4,67	4,65	<1,0	6,69	6,76	40,0
	10^{-8}	3,29	3,28	12,0	4,69	4,75	<1,0

5. Как видно из данных, приведенных в таблице, результаты математической и имитационной моделей весьма близки. Следовательно, математическую модель (1) – (5) можно использовать для ориентировочных оценок производительности соответствующих интерфейсов при заданных q, n, N , а имитационную модель – для более точных оценок применительно к заданным распределениям реальных слов БС.

Литература

1. Белоус Л.В., Литвинов В.А., Майстренко С.Я. Модель упреждающей подсказки в интерфейсе пользователя // Математические машины и системы. -2004. -№3. -С.156-163.
2. John Bonnie E. "Why GOMS ?" Interaction. October.1995.-pp 80-89.
3. A Guide to GOMS Model Usability Evaluation using GOMSL and GLEAN4 Revision: March 31, 2006, ftp://www.eecs.umich.edu/people/kieras/GOMS/GOMSL_Guide.pdf.