

# Индексирование документов, векторная модель текста

Боховко Р.Н., группа 4301

# Процесс индексирования

- ▶ Отбор индексационных терминов
- ▶ Приписывание этим терминам некоторого веса
- ▶ Отнесение каждого термина к некоторому типу
- ▶ Определение отношений

# Информация, используемая при индексировании поисковыми системами

- ▶ Позиция каждого слова
- ▶ Раздел текста, в который входит данное слово
- ▶ Шрифт и т.д.

# Понятия, используемые в векторной модели текста

- ▶ Словарь
- ▶ Поисковый образ документа
- ▶ Информационный массив

# Матрица информационного массива

	<b>Термин 1</b>	<b>Термин 2</b>	...	Термин $j$	...	Термин $D$
Документ 1	$W_{11}$	$W_{12}$	...	$W_{1j}$	...	$W_{1D}$
Документ 2	$W_{21}$	$W_{22}$	...	$W_{2j}$	...	$W_{2D}$
...	...	...	...	...	...	...
Документ $i$	$W_{i1}$	$W_{i2}$	...	$W_{ij}$	...	$W_{iD}$
...	...	...	...	...	...	...
Документ $N$	$W_{N1}$	$W_{N2}$	...	$W_{Nj}$	...	$W_{ND}$

## Список использованных источников:

- ▶ Моченов С. В. Применение статистических методов для семантического анализа текста.
- ▶ И.С. Мисуно, Д.А. Рачковский. Поиск текстовой информации с помощью векторных представлений.