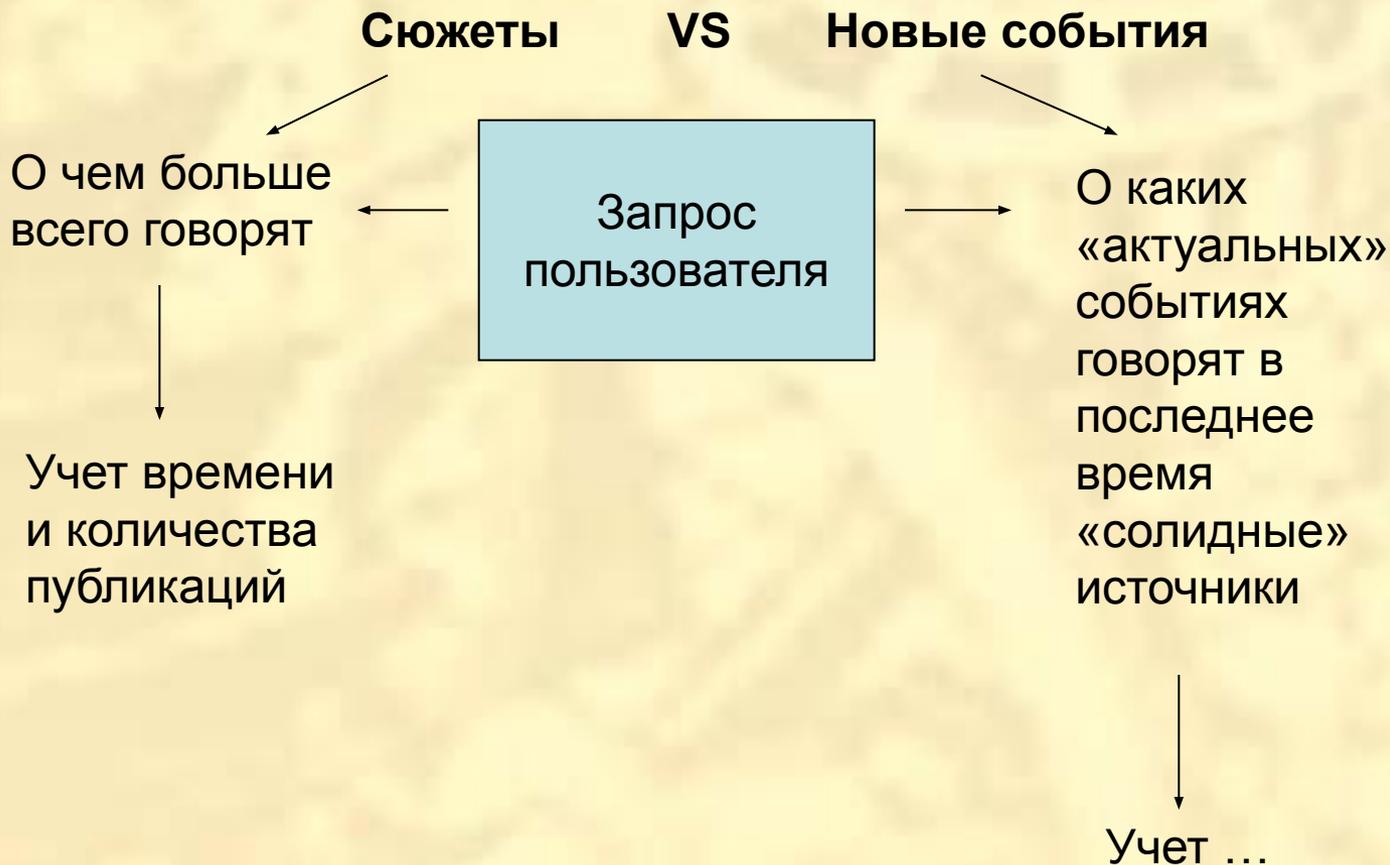


ВЫЯВЛЕНИЕ НОВЫХ СОБЫТИЙ ИЗ ПОТОКА НОВОСТЕЙ

**Ландэ Д.В., Брайчевский С.М.,
Григорьев А.Н., Дармохвал А.Т.,
Радецкий А.Б.**

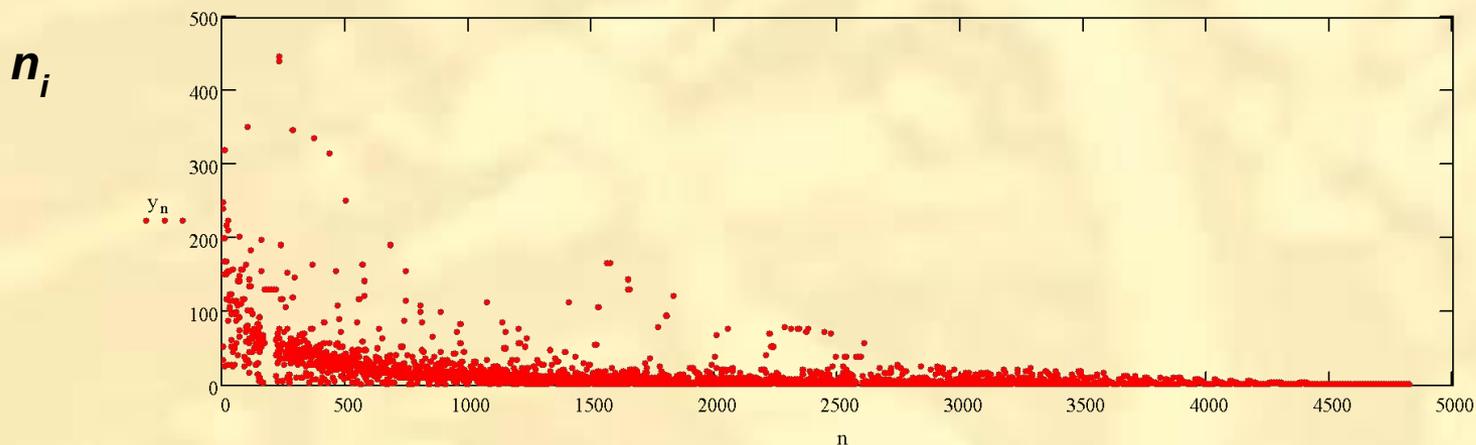
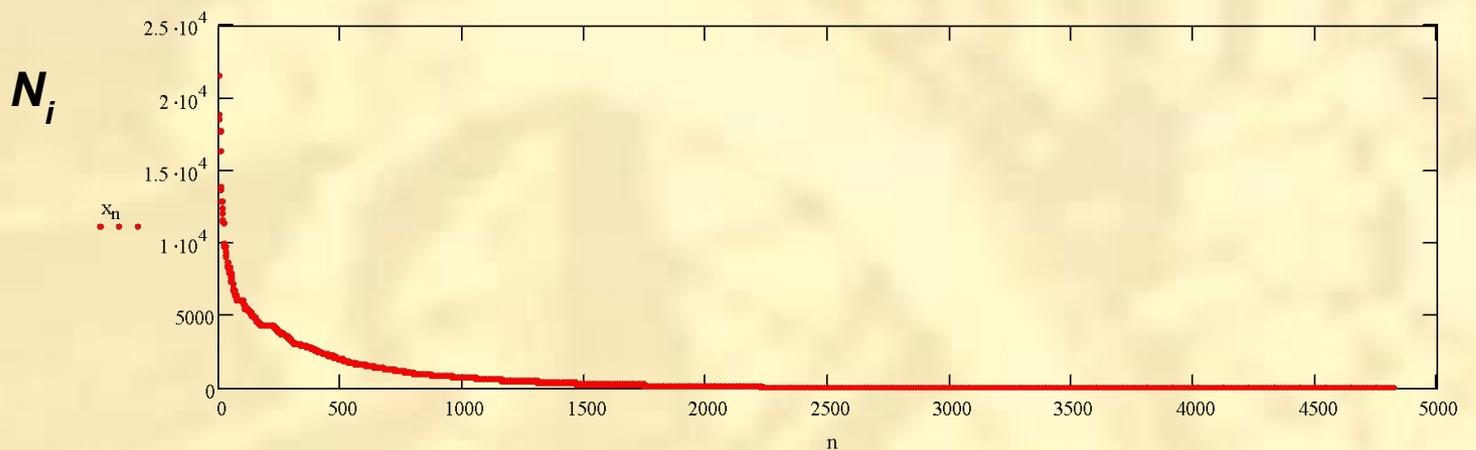
**Информационный центр «ЭЛВИСТИ»,
Киев**

ВЫЯВЛЕНИЕ НОВЫХ СОБЫТИЙ ИЗ ПОТОКА НОВОСТЕЙ



ВЫЯВЛЕНИЕ НОВЫХ СОБЫТИЙ ИЗ ПОТОКА НОВОСТЕЙ

популярность документа D_i за небольшой промежуток времени (локальная популярность - n_i) и за значительный промежуток (глобальная популярность - N_i). Вычисляются по количеству «подобных» документов в оперативной и ретроспективной базах данных.



Предположение, о том, что документ соответствует новому событию, если выполняются условия:

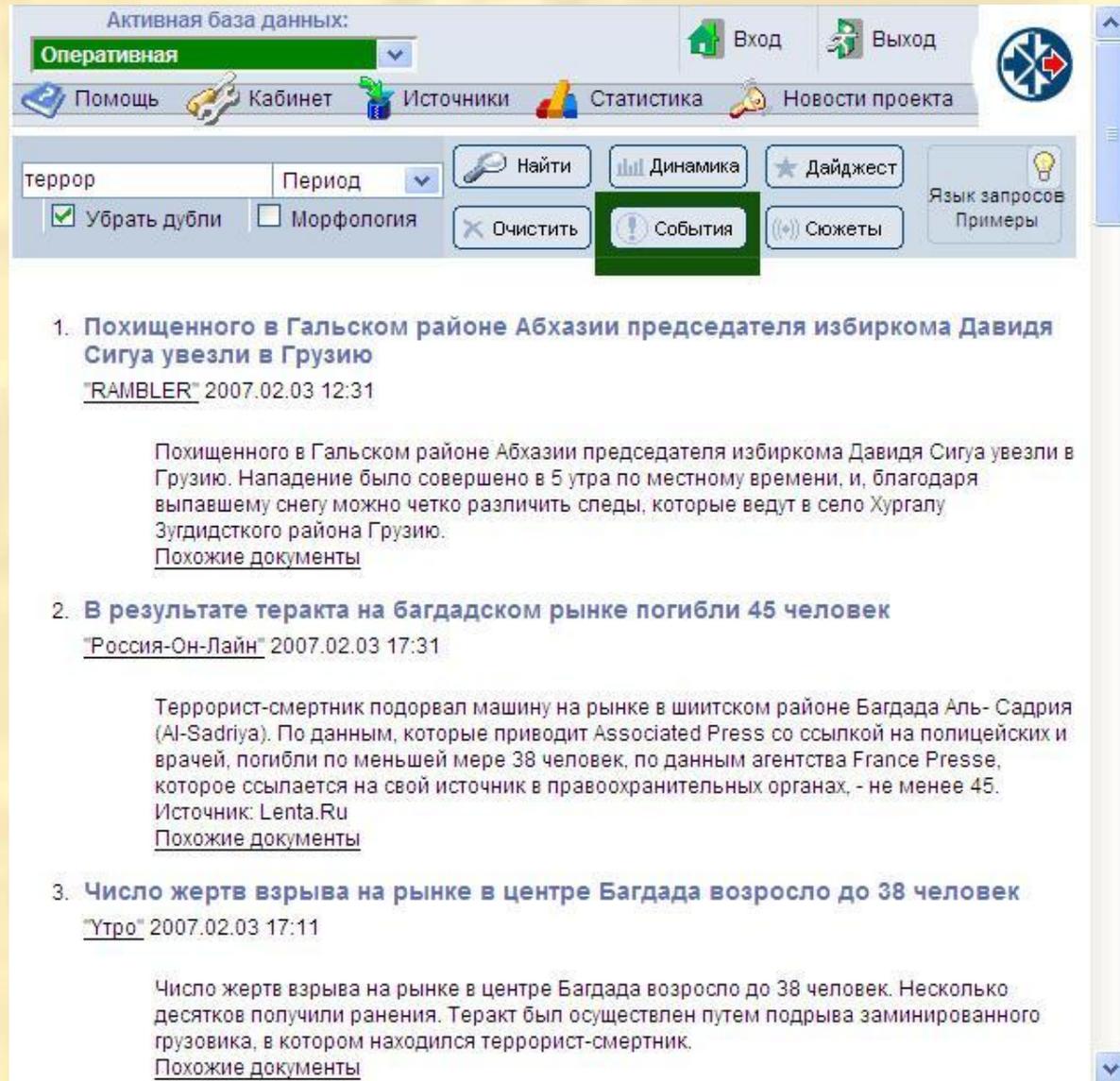
- минимальное время, прошедшее с момента публикации документа;**
- близость лексического состава документа к лексическому составу массива документов за небольшой промежуток времени (массив оперативных новостей);**
- существенное различие лексического состава документа от лексического состава массива документов за значительный период времени – окна наблюдения;**
- наличие в документе терминов, входящих в плюс-словарь (включающий важные для содержания новостей слова типа «теракт», «конфликт», «сенсация» и т.п.);**
- высокий ранг «авторитетности» источника, а также допустимости лексики заглавий новостей (определяемых экспертами).**

Формула для расчета ранга новостного документа:

$$\frac{Rang_i * sim(D_i, PlusDic) * \sum_{j=1}^n sim(D_i, D_j)}{\log(i+1) * \sum_{j=n}^N sim(D_i, D_j)}$$

- N – величина окна наблюдения потока новостей;
- n – величина массива оперативных новостей ($n < N$);
- D_i – i -й документ;
- $PlusDic$ – плюс-словарь;
- $sim(D_i, D_j)$ – мера близости документа i документу j ;
- $sim(D_i, PlusDic)$ – мера близости документа i плюс-словарю;
- $Rang_i$ – ранг источника, соответствующего i -му документу.

ВЫЯВЛЕНИЕ НОВЫХ СОБЫТИЙ ИЗ ПОТОКА НОВОСТЕЙ



Активная база данных: **Оперативная** Вход Выход

[Помощь](#) [Кабинет](#) [Источники](#) [Статистика](#) [Новости проекта](#)

террор Период Найти Динамика Дайджест Язык запросов
 Убрать дубли Морфология Очистить **События** Сюжеты Примеры

- Похищенного в Гальском районе Абхазии председателя избиркома Давида Сигуа увезли в Грузию**
"RAMBLER" 2007.02.03 12:31
Похищенного в Гальском районе Абхазии председателя избиркома Давида Сигуа увезли в Грузию. Нападение было совершено в 5 утра по местному времени, и, благодаря выпавшему снегу можно четко различить следы, которые ведут в село Хургалу Зугдидского района Грузию.
[Похожие документы](#)
- В результате теракта на багдадском рынке погибли 45 человек**
"Россия-Он-Лайн" 2007.02.03 17:31
Террорист-смертник подорвал машину на рынке в шиитском районе Багдада Аль-Садрия (Al-Sadriya). По данным, которые приводит Associated Press со ссылкой на полицейских и врачей, погибли по меньшей мере 38 человек, по данным агентства France Presse, которое ссылается на свой источник в правоохранительных органах, - не менее 45.
Источник: Lenta.Ru
[Похожие документы](#)
- Число жертв взрыва на рынке в центре Багдада возросло до 38 человек**
"Утро" 2007.02.03 17:11
Число жертв взрыва на рынке в центре Багдада возросло до 38 человек. Несколько десятков получили ранения. Теракт был осуществлен путем подрыва заминированного грузовика, в котором находился террорист-смертник.
[Похожие документы](#)

**Интерфейс
пользователя**

Спасибо за внимание!

**Ландэ Д.В., Брайчевский С.М., Григорьев А.Н.,
Дармохвал А.Т., Радецкий А.Б.**

Информационный центр «ЭЛВИСТИ», Киев

stream@visti.net