

ПРИМЕНЕНИЕ ОБЪЕКТНЫХ ФИЛЬТРОВ ПРИ ВИЗУАЛИЗАЦИИ ДИНАМИКИ ИНФОРМАЦИОННЫХ ПОТОКОВ

Ландэ Д.В., к.т.н., Григорьев А.Н., Брайчевский С.М., к.ф.-м.н., Дармохвал А.Т.,
Информационный центр «ЭЛВИСТИ»,
Снарский А.А., д.ф.-м.н., проф.,
Национальный технический университет «Киевский политехнический институт»

В настоящее время объемы и динамика обновления информации в Интернет достигли такого уровня, что позволяют говорить об информационных потоках [1]. В то же время, навигация в этих потоках является проблематичной.

Вопросы визуализации результатов поиска посвящено большое количество научных работ и практических разработок [2-4]. Вместе с тем, визуализации тенденций и объектного распределения информационных потоков больших объемов, доступных пользователям Интернет уже сегодня, не уделялось существенного внимания.

Авторами предлагается форма визуального отображения информационного потока в разрезе объектов и дат, представляющая собой таблицу (т.н., Wordlet-диаграмму), ячейками которой служат значения количества сообщений информационного потока за определенную дату, соответствующую определенному объекту. Столбцам этой таблицы соответствуют даты, а строкам – объекты, являющиеся своеобразными содержательными фильтрами исследуемого информационного потока.

В процессе построения Wordlet-диаграммы предлагается ее кластеризация путем перестановки строк (перегруппировки объектов) в соответствии с алгоритмом k-means [5]. При этом подразумевается, что например в случае разбиения на 2 кластера, основы кластеров i и j (центроиды) выбираются для наибольших значений функции:

$$F_{ij} = (\sum x_{ik} * \sum x_{jk})^{0.5} / (K_{ij} + 1)$$

где x_{ik} – значения элементов таблицы, суммы берутся по всем датам k ; K_{ij} – корреляция строк i и j .

Визуально Wordlet-диаграмма представляет собой таблицу, ячейки которой закрашены оттенками серого цвета, в зависимости от значений объемов публикации по соответствующему объекту в соответствующий день (большее значение соответствует более светлому оттенку). Следует заметить, что каждая строка Wordlet-диаграммы обладает фрактальными свойствами [6]. Фрактальные же свойства всей диаграммы являются открытым вопросом.

Авторами получены Wordlet-диаграммы, соответствующие большим потокам сообщений (свыше 50 тыс.) различной тематической направленности. В качестве объектов выбирались такие параметры, как ключевые слова, фамилии людей, бренды. На приведенной на рис. 1. диаграмме информационного потока сообщений сетевых СМИ по теме «коррупция» за 456 дней (с 1 января 2007 г. по 31 марта 2007 г.) по 30 персонам, например, отчетливо видны циклы праздничных дней, а также корреляции персон (например, персон с номерами 13 и 14 в первые 100 дней). В результате проведенных экспериментов, есть основания предположить, что использование таких средств визуализации, как Wordlet-диаграммы, позволяет «разлагать» исходные временные ряды в соответствии с объектами, обнаруживать медиа-активность по выбранным объектам, выявлять взаимосвязи объектов в разрезе дат, определять детали медиа-активности каждого

объекта или группы объектов. Использование Wordlet-диаграмм представляется важным дополнением к информационно-аналитическим системам поддержки принятия решений.

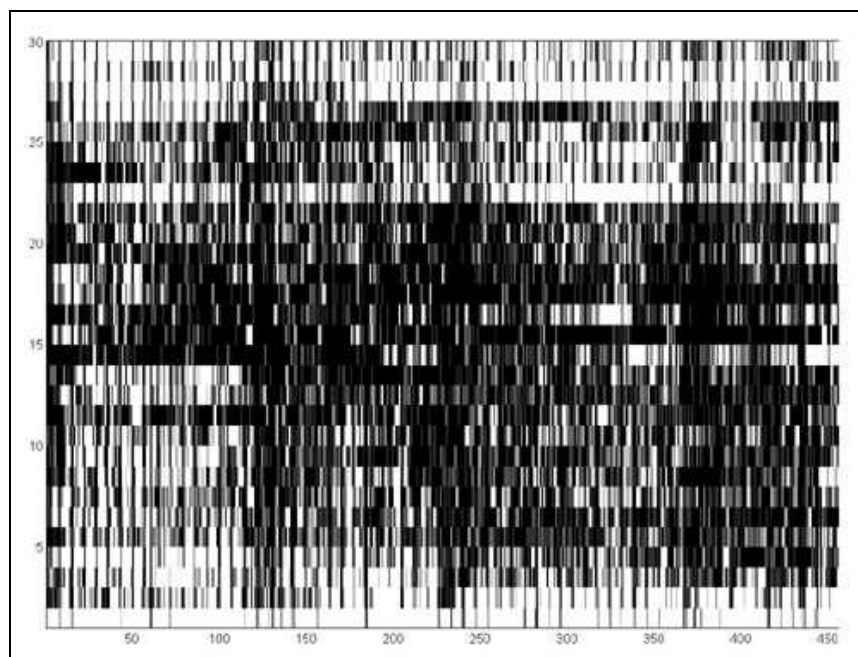


Рис.1. Пример Wordlet-диаграммы

Литература

1. Брайчевский С.М., Ландэ Д.В. Современные информационные потоки: актуальная проблематика // Научно-техническая информация. Сер. 1, 2005. - № 11. - С. 21-33.
2. M.M. Knepper, R. Killam, K.L. Fox O. Frieder. Information Retrieval and Visualization using SENTINEL / TREC 1998: 336-340.
3. Григорьев А.Н., Ландэ Д.В. Адаптивный интерфейс уточнения запросов к системе контент-мониторинга InfoStream // Компьютерная лингвистика и интеллектуальные технологии: труды международной конференции Диалог'2005 - Москва: Наука, 2005. – С. 109-111.
4. Ландэ Д.В., Григорьев А.Н. Многоуровневый классификатор-навигатор по откликам информационно-поисковой системы // Компьютерная лингвистика и интеллектуальные технологии: труды международной конференции Диалог'2006 – Москва: Наука, 2006. - С. 329-331.
5. Ландэ Д.В. Некоторые методы анализа новостных информационных потоков // Научные труды Донецкого национального технического университета. Серия: Информатика, кибернетика и вычислительная техника (ИКВТ-2005). - Вып. 93. – Донецк: ДонНТУ, 2005. - С. 277-287
6. Ландэ Д.В., Григорьев А.Н., Брайчевский С.М. Фрактальный анализ тематических информационных потоков // Доклады международной конференции «MegaLing'2006 Горизонты прикладной лингвистики и лингвистических технологий». 20-27 сентября 2006, Украина, Крым, Партенит. - С. 244-245.