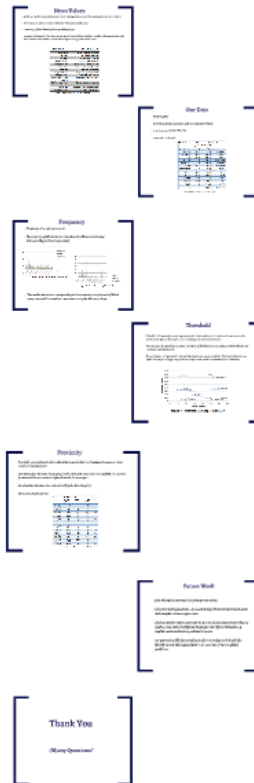


# The Pursuit of Journalistic News Values through Text Mining Techniques

Evgenia Belyaeva, Aljaž Košmerlj, Dunja Mladenčić

*Jožef Stefan Institute*

*Jožef Stefan International Postgraduate School*



# **The Pursuit of Journalistic News Values through Text Mining Techniques**

**Evgenia Belyaeva, Aljaž Košmerlj, Dunja Mladenić**

*Jožef Stefan Institute*

*Jožef Stefan International Postgraduate School*

# News Values

- *guidelines to follow in producing a news story, ideological factors in understanding decisions of journalists*
- *news values are subjective and contribute to the ubiquitous media bias*
- *a taxonomy of News Values by Galtung and Ruge (1970)*
- *our goal: to distinguish if the theory of newsworthiness by Galtung and Ruge is still a valid approach to predict news selection values and to see some interesting recurring patterns in the news*

| News Values                | Short Description             |
|----------------------------|-------------------------------|
| <i>Frequency</i>           | <i>Time span of an event</i>  |
| <i>Threshold</i>           | <i>The size of an event</i>   |
| <i>Proximity</i>           | <i>Geographical closeness</i> |
| Unambiguity                | Clarity of the meaning        |
| Meaningfulness             | Great value to the audience   |
| Consonance                 | Conventional expectations     |
| Continuity                 | Continuous over time          |
| Unexpectedness             | Unplanned/Unexpected          |
| Composition                | Other pieces of info          |
| Reference to elite nations | Relate to famous nations      |
| Reference to elite people  | Relate to famous people       |
| Negativity                 | Bad news, conflict oriented   |
| Personalisation            | Action of individuals         |

Table 2. News Values by Galtung and Ruge

# Our Data

- *Event Registry*
- *News about Apple Cooperation (iPhone 6 and Apple Watch)*
- *16 news sources in EN, DEU, SPA*
- *01.09.2014 – 31.10.2014*

| The Publisher  | Total Nr. Events | Total Nr. Articles on Apple | Headquarters  |
|----------------|------------------|-----------------------------|---------------|
| The Next Web   | 1064             | 1670                        | Amsterdam     |
| Gizmodo        | 2007             | 3911                        | New York      |
| The Guardian   | 14299            | 19997                       | London        |
| BBC            | 15582            | 23852                       | London        |
| USA Today      | 7692             | 13629                       | Tysons Corner |
| Wall Street J. | 7197             | 18837                       | New York      |
| Heise.de       | 4194             | 2190                        | Hannover      |
| Chip online.de | 907              | 1212                        | Munich        |
| Stern          | 4194             | 10092                       | Hamburg       |
| Die Zeit       | 3722             | 5600                        | Hamburg       |
| Die Welt       | 14683            | 30359                       | Berlin        |
| Der Spiegel    | 2261             | 2759                        | Hamburg       |
| El Mundo       | 6707             | 8705                        | Madrid        |
| ABC.es         | 7431             | 10388                       | Madrid        |
| El Pais        | 686              | 979                         | Madrid        |
| El Dia         | 6700             | 12752                       | Barcelona     |

Table 1. Publishers and Totals of Events/Article on Apple

# Frequency

- *Frequency- time-span of an event*
- *The frequency of all articles from the selected publishers mentioning iPhone and Apple Watch respectively*

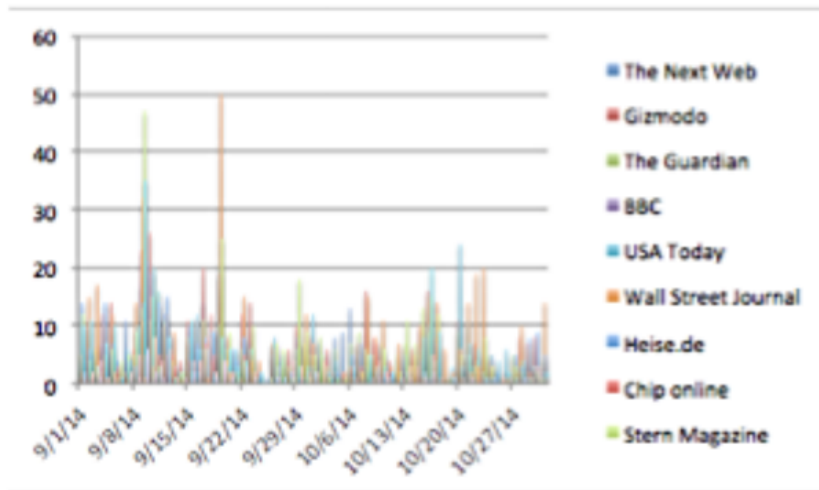


Figure 1. iPhone 6 Frequency Distribution

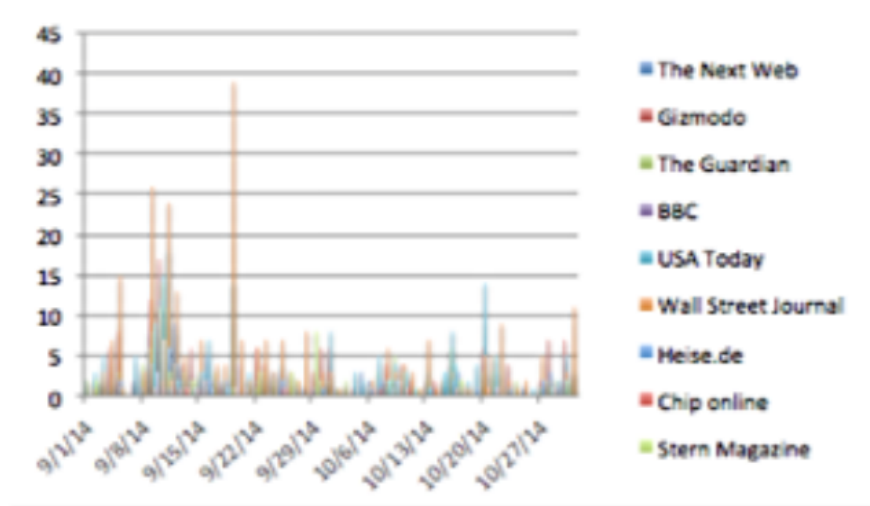
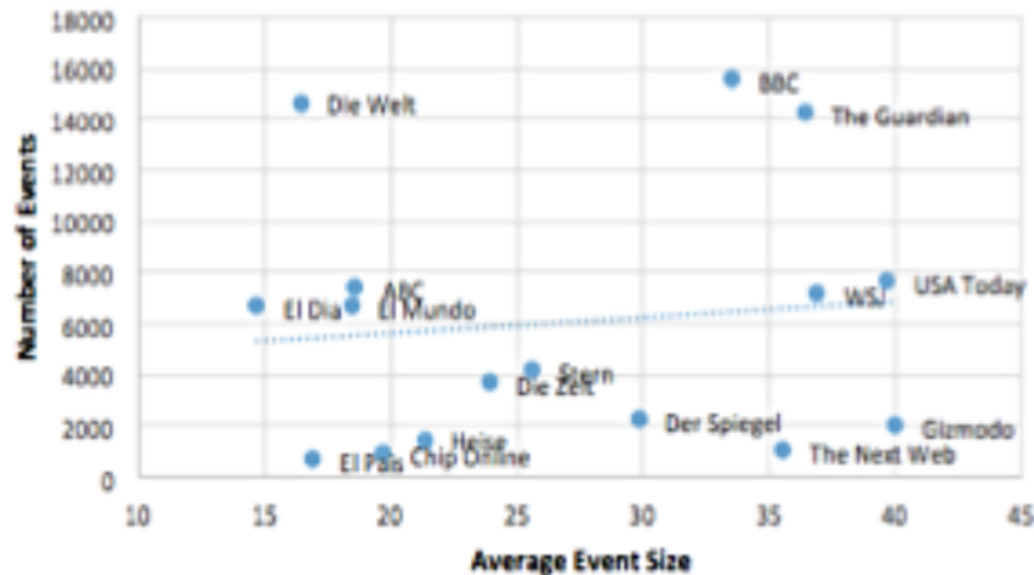


Figure 2. Watch Frequency Distribution

- *Two sudden bursts: one corresponding to the announcement of launch of Watch 09.09.2014 and the second one - announcement of the iPhone 6 release*

# Threshold

- **Threshold** - the impact of an event (a group of articles that are clustered to report on the same issues in the world) and its effect on the readers i.e. a size needed for an event to become news
- **Aim:** to capture the size of clusters (number of articles of all publishers in event clusters) without limiting our search to reports about Apple.
- **Our assumption:** a single article is not very informative, but a group of articles, which is picked up by more publishers can form a bigger story with more impact on the readers and match the threshold value.



**Figure 3. Threshold analysis per publisher**

# Proximity

- *Proximity - geographical or often cultural (in terms of religion or language) closeness of a news story to a media publisher*
- *Our assumption: the closer the geographical location of the story to the news publisher is, the more frequent and the more intense (higher threshold) the coverage is.*
- *Event location detection done automatically by the Event Registry*
- *Use of sub-selection of data*

| Publisher        | Total Nr. Country Sub-Selection | Same country | Total Nr. City Sub-Select. | Same city |
|------------------|---------------------------------|--------------|----------------------------|-----------|
| The Next Web     | 178                             | 1            | 174                        | 1         |
| <b>Gizmodo</b>   | <b>371</b>                      | <b>204</b>   | 370                        | 15        |
| Guardian         | 6563                            | 2261         | 6510                       | 462       |
| BBC              | 7105                            | 2909         | 7039                       | 438       |
| <b>USA Today</b> | <b>4299</b>                     | <b>2842</b>  | 4291                       | 0         |
| WSJ              | 3091                            | 1194         | 1074                       | 122       |
| <b>Heise</b>     | <b>586</b>                      | <b>262</b>   | 585                        | 5         |
| Chip             | 211                             | 71           | 211                        | 1         |
| <b>Stern</b>     | <b>2704</b>                     | <b>1197</b>  | 2701                       | 90        |
| <b>Die Zeit</b>  | <b>2505</b>                     | <b>1103</b>  | 2504                       | 95        |
| <b>Die Welt</b>  | <b>9185</b>                     | <b>5340</b>  | 9182                       | 1248      |
| Der Spiegel      | 1592                            | 630          | 1590                       | 55        |
| <b>El Mundo</b>  | <b>4077</b>                     | <b>2269</b>  | 4076                       | 861       |
| <b>ABC</b>       | <b>4493</b>                     | <b>2372</b>  | 4491                       | 879       |
| El Pais          | 337                             | 46           | 337                        | 7         |
| <b>El Dia</b>    | <b>3789</b>                     | <b>2399</b>  | 3785                       | 154       |

Table 4. Geographical proximity analysis per publisher



# Future Work

- *first attempt to automate detection of news values*
- *using text mining methods - an essential step of interaction between social and computer sciences approaches*
- *develop our framework: automate the process of assessing newsworthiness of all 12 news values in different languages and different domains e.g. conflicts, natural disasters, political crises etc.*
- *our approach will help journalists in their every day work, it will also identify various ideological patterns or news bias of various global publishers.*



**Thank You**

**(M)any Questions?**