

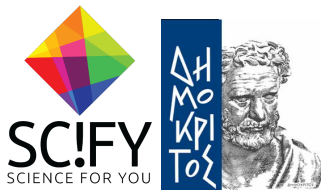
Summarization across domains and media types

George Giannakopoulos^{1,2}

¹SciFY PNPC, Greece

²NCSR Demokritos, Greece

ggianna@iit.demokritos.gr



May 2016

Why are we here?

Challenge

- Numerous text sources
- Numerous settings (social media vs. curated/edited text)
- Many languages
- Millions of users

What do we offer?

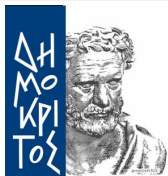
Summarization Discovering events and related documents

- Across languages
- Across media types (news, social media, websites, blogs)

Classification Routing and prioritizing information

Personalization Learning what the users like in real-time

Who are we? — NCSR “Demokritos”



- **Largest public research center** in Greece (5 institutes)
- Institute of Informatics and Telecommunications:
 - 10 labs (including 4 labs offering specialized services)
 - **Software and Knowledge Engineering Lab (SKEL)**
 - **Content Analysis and Knowledge Technologies** group
 - **Personalisation and Social Network Analysis** group

<http://www.iit.demokritos.gr/>

Who are we? — SciFY PNPC



- Private Not-for-profit company
- Brings **research results** to everyday life...
- as **open source products**...
- and **builds communities** around them.

In other words

- Exploitation
- Dissemination
- Productization
- Clear Intellectual Property Rights (IPR)

<http://www.scify.org/>

Crawling and Event detection

- Monitor
 - News feeds (RSS/ATOM); Social media; Websites; Blogs
 - Queries
 - Streams
- Identify events (as document clusters/groups)
- Record **event evolution over time**
- Identify **trends**

Summarization

- Extract main points from **all** documents



Powered by **NewSum**

Summarization

- Extract main points from **all** documents
- Avoid redundancy



Powered by **NewSum**

Summarization

- Extract main points from **all** documents
- Avoid redundancy
- Keep all event aspects



Powered by **NewSum**

Summarization

- Extract main points from **all** documents
- Avoid redundancy
- Keep all event aspects
- **Map social media texts to news events**



Powered by **NewSum**

Summarization

- Extract main points from **all** documents
- Avoid redundancy
- Keep all event aspects
- **Map social media texts to news events**
- **Discover “social-media-only” events**



Powered by **NewSum**

Summarization

- Extract main points from **all** documents
- Avoid redundancy
- Keep all event aspects
- **Map social media texts to news events**
- **Discover “social-media-only” events**
- Easily customizable on different languages (e.g. English, Greek, German, ...)



Powered by **NewSum**

Classification

- Provide customized services through class training API, i.e. make your own classification
- Can **support hierarchies**
- **Classes can evolve over time**

Personalization

- Learn what users want
- Identify market segment preferences
- Identify communities
- **Provide real-time suggestions**

Powered by



Text analytics

Combine:

- Monitoring and summarization
- Internal and external documents
- Sentiment analysis and entity extraction

IBM Hellas — NCSR “Demokritos” joint analytics services for Greek language projects

Example use cases

- Journalist tools
- Aggregators
- Topic-specific portals
- Brand monitoring
- Decision support

Why work with us?

- **Market-ready**
- Empowered by 25 years of experience in LT
- Leaders in multi-lingual summarization research
- Proven record of productization

How can we collaborate?

- Empower your services
- Build new services together
- Add value to EU projects

Reach out to us. Ask for more information.

To remember

Summarization Discovering events and related documents

- Across languages
- Across media types (news, social media, websites, blogs)

Classification Routing and prioritizing information

Personalization Learning what the users like in real-time

George Giannakopoulos (ggianna@iit.demokritos.gr)

