DeCAT 2015 - Workshop on Deep Content Analytics Techniques for Personalized and Intelligent Services

Lora Aroyo[1], Geert-Jan Houben[2], Pasquale Lops[3],
Cataldo Musto[3], and Giovanni Semeraro[3]

[1] Department of Computer Science, VU University Amsterdam, The Netherlands
[2] Delft University of Technology (TU Delft), The Netherlands
[3] Department of Computer Science, University of Bari "A. Moro", Italy

1 Introduction

According to a recent claim by IBM, 90% of the data available today have been created in the last two years. This uncontrolled and exponential growth of the online information gave new life to the research in the area of user modelling and personalization, since information about users preferences, sentiment and opinions can now be obtained by mining data gathered from many heterogeneous sources.

As an example, many recent work rely on the analysis of the content posted by people on social networks and micro-blogs to unveil latent information about their interests, automatically extract people personality traits, build preferences models on the ground of textual reviews, and so on. At the same time, the recent phenomenon of (Linked) Open Data fueled this research line by making available a huge amount of machine-readable textual data.

All these trends paved the way to the design of intelligent and personalized systems able to extract some real value from this plethora of rough textual content produced on the Web: examples of such services are online brand monitoring platforms, social recommender systems and smart cities-related applications, as incident detection systems or personalized city tour planners.

However, a complete exploitation of such textual streams requires a comprehension of the information conveyed by people. In turn, this requires a deep understanding of the language, which is not trivial. The major goal of this workshop is to stimulate the attention of the scientific community on the aforementioned topics. The workshop aims to provide a forum for discussing open problems, challenges and innovative research approaches in the area, in order to investigate whether the adoption of techniques for semantic content representation and deep content analytics can be effective to build a new generation of intelligent and personalized services based on the analysis of Social, Big and Linked Open Data.

2 Motivations and Workshop Topics

The importance of user modeling and personalization is taken for granted in several scenarios. According to this widespread paradigm, each user can be modeled to some (explicitly or implicitly gathered) information about her knowledge or about her preferences, in order to adapt the behavior of a generic intelligent system to her specific characteristics.

However, the rapid growth of social networks changed the rules for personalization, since the spread of these platforms radically changed and renewed many consolidated behavioral paradigms. Indeed, people today exploit these platforms for decision-making related tasks, to support causes, to provide their circles with recommendations or even to express opinions and discuss about the city or the place where they live. Thanks to the heterogeneous nature of the discussions that take place on social networks, a lot of new data are continuously available and can be gathered and exploited to build richer and more complete user models, to discover latent communities, to infer information about users emotions and personality traits, and also to study very complex phenomena, such as those related to the psycho-social sphere, in a totally new way. At the same time, thanks to crowdsourcing, a huge amount of content-based information has been made available in open knowledge sources as Wikipedia and the Linked Open Data Cloud.

Given that most of the information stored in these modern data sylos is made available as textual content, a consequence, a complete exploitation of these rich information sources requires a big effort on the definition of models and techniques able to effectively process the content and to represent it in a machine-readable form, in order to unveil the latent semantics and trigger more effective personalization and adaptation pipelines. This is not a trivial task, since this process requires a deep comprehension of the language, which in turn typically requires a combination of techniques coming from Machine Learning and Natural Language Processing areas.

The main goal of the workshop is to stimulate the discussion around problems, challenges and research directions regarding the exploitation of content-based information sources (Big, Social and Linked Data) for personalization and adaptation task and to foster the design of a new generation of intelligent user-centered services.

We hope the workshop will stimulate discussions around the presented papers, the invited talk and the following questions:

– What is the impact of semantics in personalization and adaptation tasks?
– Can social media improve the representation of user interests?
– Can semantic analysis technique improve the representation of user interests?
– Can these data sylos (Wikipedia, DBpedia, Freebase) be useful for personalization and adaptation tasks?
– Which data sylos are more effective to model user interests and preferences?
– What content-based information is more useful to personalize and adapt the behavior of modern intelligent systems?

- Does a semantic representation of the information improve the effectiveness of personalization tasks?
- Does a semantic representation of the information improve the transparency of such platforms?
- Can the analysis of content coming from social media provide some information about user personality traits?
- How do people deal with privacy issues? Are them willing to trade better personalization with a larger tracking of their activities on the Web?
- Is it possible to think about a novel generation of adaptive platforms able to completely exploit all the available information?

3   Contributions

Five papers will be presented in DeCAT 2015. The papers were accepted after a peer-review process: each paper was reviewed by at least two members of the Program Committee and evaluated in terms of Significance, Technical Quality and Novelty of the approach.

In their contributions, Abela et al. [1] tackle the Personal Information Management (PIM) problem, and propose a methodology to automatically organise personal information accessed by the user into task-clusters. To this aim, the authors transparently exploiting the users behaviour while performing some tasks. A distinguishing aspect of their work is the usage of PiMx app. a tool which can be of interest for other researchers working on task clustering.

Next, Papadopoulos et al. [2] present ongoing work on the formalization of a persons creativity, modelling it in terms of four characteristics of the personal content creations, namely novelty, surprise, rarity and recreational effort. Based on such formalization, the paper also presents the Creativity Profiling Server (CPS), a system implementing the aforementioned user modelling framework for computing and maintaining creativity profiles

The analysis of social media is the focus of the work proposed by Matta et al. [3]. In this paper the authors perform an interesting analysis of the connection between Bitcoin's price and the volume of Tweets about the topic. Specifically, the authors use an external API to crawl Twitter data and assign a sentiment to it. Next, they analyze how the price of Bitcoins changed over time and they looked for some connections between these aspects. A thorough analysis of the time series showed that some connection (calculated as the cross-correlation between time series) exists.

In the only short paper accepted, Pentel investigated the relation between reading and writing skills in the task of age-based categorization. In this contribution [4] he presents results of a study on age-based categorization of short texts as 85 words per author. He introduced a novel set of features that will reliably work with short texts, which makes easy to extract from the text itself without any outside databases.

Finally, Basile et al. [5] propose a content-based and time-aware movie recommendation approach. The novel contribution is the time-adaptivity for a content-based technique. The authors proposed an approach that models short-term

preferences by adopting a content-based sliding window approach: when a new ratings comes into the system, the replacement of an older one is performed by taking into account both a decay function for user interests and content similarity between items on which ratings are provided, computed by distributional semantics models. The authors carried out an evaluation that demonstrate that their approach overtake the baseline FIFO strategy.

References

1. Abela, C., Staff, C., Handschuh, S. *Automatic Task-Cluster Generation based on Document Switching and Revisitation.* In Proceedings of DeCAT 2015 - 1st Workshop on Deep Content Analytics Techniques for Personalized and Intelligent Services, co-located with UMAP 2015, Dublin (2015).
2. Papadopoulos, G., Karampiperis, P., Koukourikos, A., Konstantinidis, S. *Creativity Profiling Server: Modelling the Principal Components of Human Creativity over Texts.* In Proceedings of DeCAT 2015 - 1st Workshop on Deep Content Analytics Techniques for Personalized and Intelligent Services, co-located with UMAP 2015, Dublin (2015).
3. Matta, M., Lunesu, M.I., Marchesi, M. *Bitcoin Spread Prediction Using Social And Web Search Media.* In Proceedings of DeCAT 2015 - 1st Workshop on Deep Content Analytics Techniques for Personalized and Intelligent Services, co-located with UMAP 2015, Dublin (2015).
4. Pentel, A. *Employing Relation Between Reading and Writing Skills on Age Based Categorization of Short Estonian Texts.* In Proceedings of DeCAT 2015 - 1st Workshop on Deep Content Analytics Techniques for Personalized and Intelligent Services, co-located with UMAP 2015, Dublin (2015).
5. Basile, P., Caputo, A., de Gemmis, M., Lops, P., Semeraro, G. *Modeling Short-Term Preferences in Time-Aware Recommender Systems.* In Proceedings of DeCAT 2015 - 1st Workshop on Deep Content Analytics Techniques for Personalized and Intelligent Services, co-located with UMAP 2015, Dublin (2015).