Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 75 (2015) 66 – 74

Procedia Computer Science

2015 International Conference on Virtual and Augmented Reality in Education

Best-Practice Piloting Based on an Integrated Social Media Analysis and Visualization for E-Participation Simulation in Cities

Dirk Burkhardt[a,b,]*, Kawa Nazemi[a], Egils Ginters[c]

[a]*Fraunhofer IGD, Fraunhoferstr. 5, 64283 Darmstadt, Germany*
[b]*TU Darmstadt, Fraunhoferstr. 5, 64283 Darmstadt, Germany*
[c]*Sociotechnical Systems Engineering Institute Vidzeme University of Applied Sciences, Cesu Street 4, LV- 4201 Valmiera, Latvia*

Abstract

E-Participation and the engagement of citizens in politics are on the agenda in many countries. Therefore, a number of project and programs were initiated that aim at finding creative and effective solutions for involving citizens. One intended aspect is the provision of transparency that enables citizens to analyze socio-economical indicators and serves as basement for initial data storing for e-participation scenarios simulation. A beyond going step is to consider citizens' opinions in the policy making process. Unfortunately this will only be possible; if a significant number of citizens know what concrete aspect should be realized. The growth and establishment of social media nowadays allow most people to discuss political ideas and critics virtually and anonymously achieving higher credibility of the data usable for policy making. However, the major barrier is the different topical languages, the representatives' role and weak visualization of the results. This article dealt with a best-practice piloting approach for discussing and realizing innovative solutions with a variety of stakeholders from different domains. The core of this paper is how to deal and manage the different perspective on development process, to aim on designing a common set of requirements and development procedure. The methodology is practically and beneficial applied as social media integration solution in the EC FP7 FUPOL research project.

Peer-review under responsibility of organizing committee of the 2015 International Conference on Virtual and Augmented Reality in Education (VARE 2015)

Keywords: Policy making and simulation; semantic visualization; e-participation.

* Corresponding author. Tel.: +49-6151-155-578; fax: +49-6151-155-139.
E-mail address: Dirk.Burkhardt@igd.fraunhofer.de

1877-0509
Peer-review under responsibility of organizing committee of the 2015 International Conference on Virtual and Augmented Reality in Education (VARE 2015)
doi:10.1016/j.procs.2015.12.214

1. Introduction

E-Participation and the engagement of citizens in politics is on the agenda in many countries. Therefore, a number of project and programs were initiated that aim to involve citizens. One aspect is the provision of transparency, e.g. through Open Data provision that enables citizens to analyze indicators of a country, state, province or the local region. A beyond going step is to consideration of citizens opinions in the policy making. In many countries petitions are one of the option to bring an important issue (from the citizens' point of view) on the political agenda. Unfortunately this will only be possible, if a significant number of citizens know what concrete aspect should be realized. Even more, it is not realistic that each issue, in particular on municipality level, can be clarified by the citizens via petition.

A modern approach to recognize opinions of citizens is by using ICT. The growth and establishment of social media allow most many people to discuss political ideas and critics virtually. However the implementation of e-participation scenarios asking for procedures assessment and convenient visualization of the results.

2. A new integrated approach for social media analysis and visualisation

To face the challenge of a novel social media analysis approach and the required sustainability for that public authorities need to work with ICT. It consist on three major technologies that work together: the FUPOL Core Platform[1], the Hot Topic Sensing and the visualization as major user-interface the user work with. Additional part is e-scenarios simulation giving forecasts about expectable results and about credibility of those.

2.1. Overall structure and architecture

The major difference to most other available social media processing systems is the modular structure (see Fig. 1). The whole system consists of several modules or rather plug-ins that can be replaced or extended by new modules. Therefore, it is not a complete proprietary system that is limit on the functions a single enterprise has implemented. Even more, the APIs allow to connect modules that other organization do provide who have more experiences in a certain domain, e.g. in visualizations. The result of such an open architecture approach is the ability to implement the most beneficial modules for a certain task and having therefore more productive system the user has to work with.

2.2. Crawling the social media content

The first step of processing social media data is the crawling. For this purpose the user has previously defined a so called campaign and to such a campaign he aligned a number of social media channels. A social media channel is for instance a specific user page, a social media page or an RSS feed. The idea is therefore a specific observation of certain sources. This is essential because especially on city level there are just a small number of well-known virtual discussion places of political interested citizens, as for instance the twitter channel of local opinion leader or discussions of citizens on digital articles of local newspaper websites. However during the last years a new problem appears. Unfortunately nowadays some special funded armies of commentators operating in social media environments. Their task is to fight with propounded viewpoint of opposite army and push into media their declared policy viewpoint. Therefore sometimes biggest part of comments are related with propaganda neither are real opinions of the citizens' about issued policy activity.

To crawl the content, two types of connectors are distinguished. The generic and the specific connectors. The generic connectors are ones that can be used for multiple sources, as for instance an RSS connector. With this kind of connector various kinds of sources, which is often provided on websites, can be crawled without making changes on the connector. The disadvantage of this type of crawler is the limitation. The crawled source depends strong on what data is provided via RSS, this can include, e.g. the title, description, back links and comments of an article, but is has not. Even more many newspaper pages reduce the text on the first few sentences. To get the full range of available data the other group, i.e. the specific connectors can be used. The task of such a specific connector is to get as much data as available from a single data source. Such a specific connector can be used for twitter, facebook or for a certain blog engine. Such connectors are often able to the full text content, and even more meta-information as for instance corresponding likes, shares and comments. Because of the fact the specific connectors are time consuming (also later when the connectors needs to be maintained, because the API of data source can change over time), it is to weight for what data source this effort is beneficial and for which the generic connectors are adequate enough.

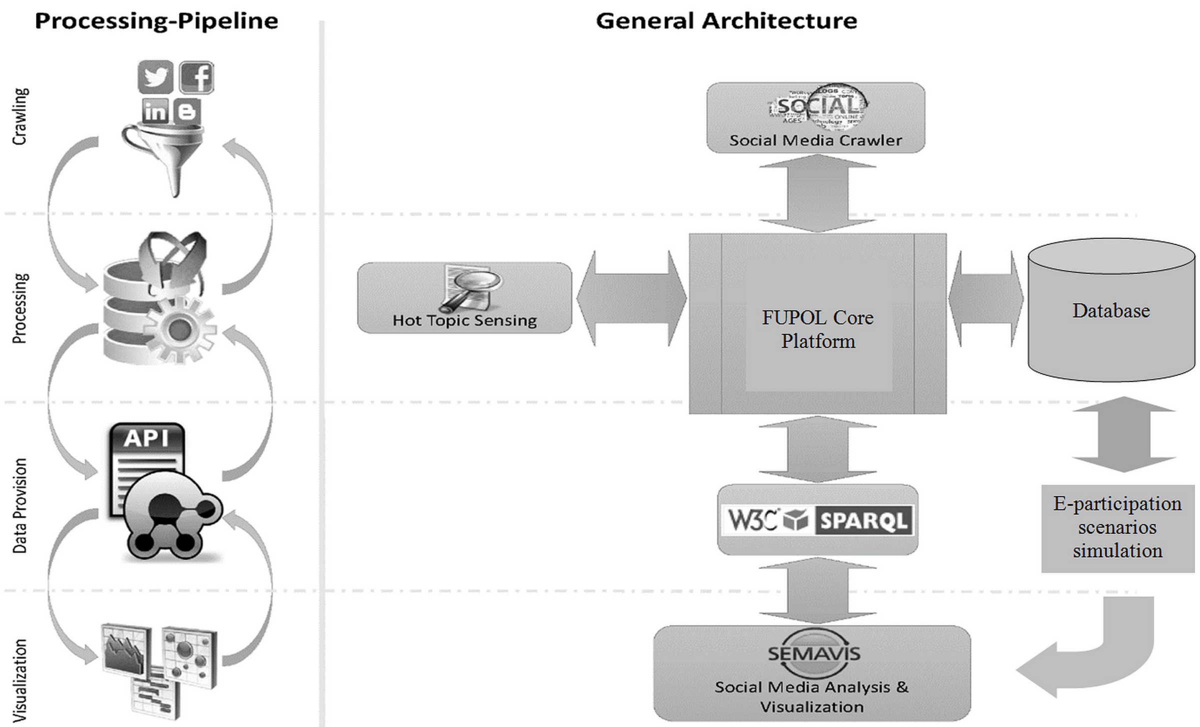

Fig. 1. Architecture of the visualization interaction among social media data.

As result of the crawling phase postings from the mentioned sources were crawled and cached for the following text mining. Next to the specific contents, also additional information are stored, as for instance the number of likes, written comments etc. Critical data, e.g. the user names and profiles, are anonymized to ensure data privacy.

2.3. Data processing and management

The data processing consist currently off three steps. The first step is the hot topic sensing. Based on all the available postings to a certain campaign, topics were automatically generated by the use of LDA algorithm[2]. These topics representing a kind of categorization to allow later a filtering based on the purpose of an analysts (urban planning, waste management, public relation). Because of the fact that LDA algorithm generating topics just based on extracted keywords, the quality of the resulting topics are less human readable. To provide a better readability, the system can be improved by the definition so called categories. A category is aggregation of many topics, which the user has to align manually. The result is in best case a hierarchy of human understandable categories to which topics are aligned and representing a bunch of postings.

The second step is semantification of the available data. To structure the data, an owned defined ontology is used, but with usage of already existing namespace definitions. It considers a number of features - among others, the categories, the topics, the postings and the authors.

The third step is the persistence of the generated data in a relational database. Based on the basic features of the ontology, all entities of the ontology are stored in the database in relational schema.

2.4. E-participation scenarios simulation

The initial data and further simulation results are stored in POSTGRE SQL database. Collected data allow to policy makers running and checking different what-if situations on the basis of agent-based simulation models, e.g. recreational resources occupancy analysis, regional economic development and justified bicycle routes and intermodality planning[3]. For convenient visualization of the simulation results Social media visualization and analysis module can be used.

2.5. Social media visualization and analysis

The main challenge of visualizing the social data is the masses of instances in the described semantic representation. Two ideas of technologies were elaborated to face this problem on the data level, but beside a solution reducing the amount of instances per class/concept, the challenge of visualizing a mass amount of data still remains. An adequate way of facing this challenge on the visualization-level is the appliance of Shneiderman's Information Seeking Mantra. Shneiderman proposed a three-level seeking mantra[4] containing the following steps: overview first, zoom and filter then details-on-demand. In the context of visualizing the social information the overview aspect plays a key role. In particular, related with the context of social data visualization two main views on this information-level were identified:

- On categorical level
- On temporal level

The levels of overview visualizations are not distinct and can be combined to view on different information aspects.

The thematic arrangement enables a visual overview definition of "categories-of-interest", whereas all are some part of information are visualized interactively. In this context two main visualization types are applied to visualize the relevance computed and the result of a quantitative analysis on the user request. The different informational requirements are then visualized on the presentation level by using the visual variables. The size of a graphical entity will provide quantitative information whereas the relevance is visualized by their color.

As the first categorical visualization so called ThemeRiver[5] is provided, which visualizes the topics and the weights over time. This visualization addresses both of the above mentioned information levels: categorical and temporal.

At this stage, the user can analyze upcoming relevant topics as well as important topics. By selecting a topic, the user filters the data in significantly on a special part. With visualizing the temporal overview and providing a faceting in time another dimension of the data is investigated. The temporal view is the most beneficial way to:

- View the trend of upcoming social opinions
- Interacting with and filtering semantic data for topic-of-relevance based on time

Here, the use of a stacked graph is proposed with the using the following informational requirements on the information dimensions:

- Size: quantity of topics, terms or extracted features
- Color: relevance based on the computed relevance
- X-Axis: temporal spread

As the second categorical visualization an hierarchical treemap is provided that uses the thematic hierarchy of the ontology as one visual indicator, the relevance of the topics as another visual indicator and the size as a third indicator for providing an overview of a topic on categorical level. The parameters are abstracted to highest level. The hierarchy is simplified visualized as an overlapping (superimposing) and integrating spatial spaces. The size is illustrating the quantity and the color the relevance (see Fig. 2).

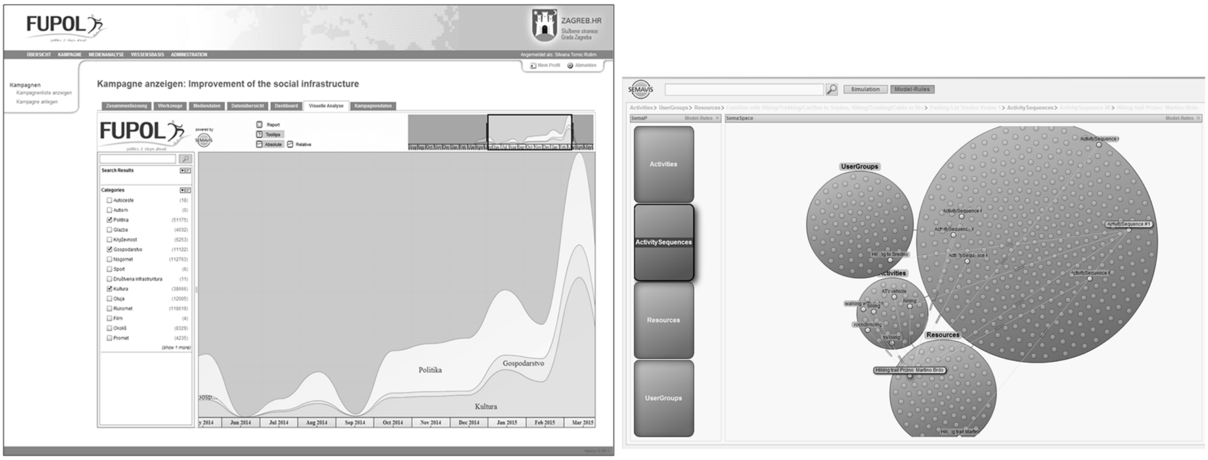

Fig. 2. Semantic data visualization views.

In contrast to that very simple visual view, a graph-based layout will be integrated that targets on the same information values. Therefore the size of circle will be used as the indicator for the quantity of information in one category, the hierarchy will be displayed as smaller integrated circles, and the color will be used for the computed relevance. Any semantic relationships in this view is dismissed, to not confuse the user with too many information.

Next to the categorical filtering, also mechanism for a temporal filtering are provided. For this purpose the timeline visualization SemaTime[6] is applied. In regards of the performance limitation of the SparQL-endpoint at the beginning of the Zooming & Filtering stage, postings on a post number lower than 50 pieces are produced. Through an enlargement of the timeframe at the bottom of the SemaTime, the time range can be increased and in the follow further postings will be retrieved by the SparQL[7]-endpoint.

The next step after the overview is a more detailed view with relational information. Therefore the existing graph-based visualizations will be extended to visualize the dependencies between actors and topics, between actors themselves and between topics themselves. This step can be done after a refinement on the overview visualization or based on a specific search that contains a comprehensible number of entities.

A force-directed visual graph algorithm with quantitative analysis for this issue is used. In this case the size of a circle indicates the number of entities, the color the relevance, the size of entities the number and/or relevance of a topic or actor himself and the relations the semantic relationship design in the FUPOL social data ontology.

The detailed visualizations can further provide more information by requesting more details on demand. For example in the figure, one actor would have a greater size than the others. With this information it can be assumed that this actor is an opinion maker, because either he has many postings or the postings are read by many people (regarding to the underlying data and goal). By clicking on this actor the visual representation will first give more information about him and further provide detailed information (as far as available) about the person.

In all the steps different visualization types are defined that are appropriate to meet the informational requirements from the social data part of view, described in D5.2[8]. One of the main contributions in this task is that the visual change of the steps from overview to details and vice versa is recognized and appropriate visualizations are provided in combined user interfaces.

To provide visualization of e-participation scenarios simulation results SemaVis[9] is used (see Fig. 3). The SemaVis framework is designed to visualize semantic information by offering effective navigation and interaction mechanisms. A special feature consists in the adequate visualization for different user groups with different preferences and background knowledge, both in terms of information to be displayed as well as in interacting with the visualizations. These users and user group orientation with customizable look and feel of principles had been developed and integrated into the SemaVis framework.

Fig. 3 shows an example of SemaVis application integrated with Skopje Vodno Mountain Recreational Activities simulator[10]. SemaVis application screen consists of several (manageable) blocks. One of the blocks contains an integrated simulator with simulation results; other blocks around it contains different visualization for this particular simulation. User can add new visualization approaches, choosing from the right side menu. In this example (see Fig. 3) there are four data representation blocks, besides simulator block, data structure block, data structure as graph block, raw data table block, line chart block and candlestick chart. All of these blocks are linked to simulator block.

Fig. 3. Vodno Mountain simulator integration with SemaVis[10].

The approach mentioned above ensures interactive visualization of simulation results as statistical data. The visual analysis of the results can be performed: drilling-down navigation to select the relevant indicators; identifying of relevant influencing factors or impacts; interlinking of various data-sources for an advanced information acquisition and context comprehension; combining of an analysis cockpit to see the data from “different perspectives”.

3. Piloting and establishing social media analysis and e-participation simulation in cities

The deployment of new ICT in public authorities is a challenging point, because of high requirements regarding stability, usability and expected benefit of a new system. One of the reason is that public authorities need systems that do not limit the productivity, since they have to follow strict workflows. To consider these specific behaviors and requirements, it was essential to stay in contact with those experts and define and clear procedure how and what aspect are necessary. Besides that the simulators have difficult and specific interface, which is based on opinions of local authorities and policy makers. It is very challenging to provide similar simulators for different audiences[11].

3.1. Finding a common language: Technicians vs. Public authority policy makers

The very first one of the most important requirement was the finding of a common language. This aspect sounds easier as it really is. Researchers and technicians have mostly a limited view speech that orients on the technical behavior. They less consider the political aspect only in a limited way (this should not mean that these aspects are neglected overall). On the other hand the employers of public authorities have almost a less technical background, in particular in regards of research and development. They know there tasks and what they currently use for tools to solve these tasks. To define a common development roadmap it is essential to understand the conditions of both sides. Therefore it is essential to communicate the conditions to other side.

To achieve such a common language, it is not possible to mention a concrete solution, because in any case an individual solution must be found based on discussion. In general it is important that both side see the opportunities of such a cooperation, so that “the good will” is available on both sides.

3.2. Multi-user requirements analysis

If the will is available on both sides and also a common understanding and language is found, it is essential to define a roadmap how to get from visions to a sketch and finally to a solution. In FUPOL an own defined requirement analysis process for the Use-Case Requirements (UCR) analysis was used, which covers all relevant parts to find a common agreement and at the end a successful solution. The procedure comprises: domain identification, elicitation (categorization, illustration), abstraction (sorting, grouping), specification (description, allocation), review (verification, illustration) and negotiation (consultation, validation, supplement/amendment) (see Fig. 4).

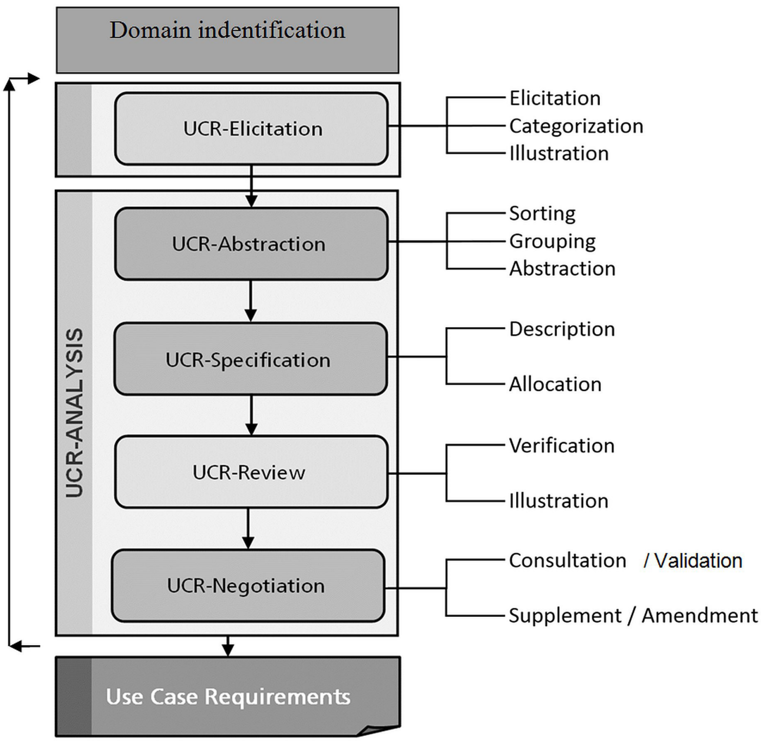

Fig. 4. Hierarchical model of use-case requirements analysis.

Overall this model can be seen as an iterative approach. After the first round of the UCR analysis and implementation of discussed features, another round can be initiated to refine and extend the system. This makes this approach useful to use also in large-scale projects as for instance in FUPOL or in general in bigger projects where more than just one development iteration is intended.

One important fact is that this approach should be performed before major developments in perspective of user related features are in progress[11]. Basic aspects that are relevant overall, like the development of a general system body, a configuration parser etc. can be developed in anyway parallel to the UCR analysis. But in particular the graphical user-interface and in regards of that some data-processing aspects should be shifted at the end of the UCR analysis.

3.3. Engaging meanwhile software development

Many software development cycles follow the approach of first defining the requirements and goals. In the second phase the system are developed and only after detailed validation it will be deployed and provided to the users. A major challenge of this approach is that misunderstandings and wrong implemented features can be identified before the whole development is done.

To keep communications between providers and beneficiaries as most efficient as possible, it is important the both sides understanding that this is not a final software and only the relevant results should be discussed. In particular the graphical user-interface is often liked to discuss in detailed, also if the used layouts and styles are not completed. Therefore it is important to focus on the general approach of the user interface and it fits to the tasks and expectation of the beneficiaries. So, the general interaction metaphors and styles during the development should be more in scope than graphical aspects.

3.4. Smart feedback collection for technology improvements

To collect empirical information about general used metaphors or selection visualization algorithms, these approaches mentioned above are limited or only give a subjective impression. In particular if graphical visualizations are used, it will be hard to identify if the visualization really provide an added value and allow a more effective and efficient work. The subjective feedback of users are good orientation, but it is more pragmatic and empirical to validate this based on analysis of e.g. task completions or the used task completion time.

Currently there are only a rare number of approaches established that allow such analysis aspects in efficient manner. Almost all approaches dealing with local methods and do not allow to provide an evaluation via web as distances-based evaluation. For this purpose an evaluation approach is designed to make such a distance-based evaluation over web and collecting empirical data based on real tasks the participants (target users) have to solve.

In evaluation system it is possible to use classic questionnaires, e.g. to collect demographic information about the users as well as further information about his ICT skills. More important is the novel approach to evaluate a system in an empirical manner. Hence, the user has also to deal with the novel part of our evaluation system, he get a number of tasks that he should answer with the system. In the background the evaluation systems logs task completion, the task completion time and information, on which a detailed analysis can be performed to identify gaps and aspects that should be enhanced.

In a combined view of the conventional questionnaires and the practical evaluation, it is possible to identify if the subjective perception of users is different to the objective facts that is measured practically.

4. Conclusion and future work

The conception about goal system functionality will always differ between the policy decision maker and the developer because the decision maker is guided by specific functionality requirements, nuances and particular conditions only known to him, whereas the developer places importance on technical solutions, development time and potential costs. The complexity of tools to be developed increases not only from a technical (territorially distributed systems, cloud computing, Future Internet architectures, semantic search etc.), but also from a functional standpoint (decision-making algorithms; fusion, mining and useable visualisation of complex data). This significantly complicates the development of suitable software. one of the ways how to implement successful projects is social inclusion of the citizens' and policy makers through e-participation.

Other important aspect is convenient visualization of the results in conformity with domain understanding. The categorical, temporal, and in future the geographical view of the data can be combined in various ways to provide a sufficient view on the social data. One promising way to provide a fruitful way for visualizing the different informational requirements of social data and statistical data respectively is the juxtaposed orchestration of visualizations.

It is clear that quality of the validation of designed technologies depends on received feedbacks. Interviewing has low credibility and efficiency. Future is into semantic search and analysis tools development allowing qualitative and intelligent gathering of the data in social media.

Also close is the time when Future Internet will be introduced. Unfortunately software engineers working on the concepts and software are not familiar with some specific fields of services like visualization, virtualization and simulation, and mostly thinking in typical SoA and web-services categories, ignoring special requirements for the message length and frequency. So, therefore close is the time when all the mistakes already done during Future Internet designing will be clear and the specialists who must deploy a new application services on the net will have a huge job.

Acknowledgements

The article publication is initiated by FP7 FUPOL project No. 287119 “Future Policy Modeling".

References

1. FUPOL. Politics 2 Steps Ahead. Retrieved: 01.06.2015, URL: http://www.fupol.eu/en
2. Blei DM, Andrew Y, Jordan M. Latent Dirichlet allocation. In: Lafferty J, editors. *Journal of Machine Learning Research* 2003; **3**(4–5): 993–1022.
3. Ginters E, Aizstrauts A, Dreija G, Ablazevica M, Stepucev S, Sakne I, Baltruks M, Piera MA, Buil R, Gusev M, Velkoski G. Skopje Bicycle Inter-modality Simulator – E-involvement through Simulation and Ticketing. In: *Proceedings of 26th European Modelling & Simulation Symposium (EMSS 2014)*, 2014. Bordeaux, France, pp. 557-563.
4. Shneiderman B. The Eyes Have It: A Task By Data Type Taxonomy For Information Visualization. In *Proceedings of IEEE Visual Languages*; 1996. pp. 336-343.
5. Ware C. *Information Visualization: Perception for Design*. 3rd ed. Elsevier; 2012.
6. Nazemi K, Stab C, Kuijper A. A Reference Model for Adaptive Visualization Systems. In: *Human-Computer Interaction. Design and Development Approaches. Lecture Notes in Computer Science;* **6761**; 2011. p. 480-489.
7. Haag F, Lohmann S, Ertl T. SparqlFilterFlow: SPARQL Query Composition for Everyone. In: *The Semantic Web: ESWC 2014 Satellite Events. Lecture Notes in Computer Science*; **8798**; 2014. p. 362-367.
8. Burkhardt D, Nazemi K. Deliverable 5.2–Visualization Software Design Description. 2013. Available on: http://cordis.europa.eu/project/rcn/100737_en.html
9. Nazemi K, Breyer M, Burkhardt D, Stab C, Kohlhammer J. SemaVis - A New Approach for Visualizing Semantic Information. *In Towards the Internet of Services: The Theseus Program*, 2014, pp.191-202. Springer International Publishing.
10. Ginters E, Aizstrauts A, Baltruks M, Nazemi K, Burkhardt D, Sonntagbauer P, Sonntagbaur S, Gutierrez JM. FUPOL Simulators and Advanced Visualization Framework Integration. In: *Proceedings of 26th European Modelling & Simulation Symposium (EMSS 2014)*, 2014. Bordeaux, France, pp. 523-530.
11. Ginters E, Aizstrauts A, Aguilar-Chinea RM. Sociotechnical Aspects of Policy Simulation. In: Sonntagbauer P, Nazemi K, Sonntagbauer S, Prister G, Burkhardt D, editors. *Handbook of Research on Advanced ICT Integration for Governance and Policy Modeling*. IGI-Global Publishing; 2014. p. 113-128.