Requirements and Architecture of a Passive Crowdsourcing Environment

Yannis Charalabidis[1], Vangelis Karkaletsis[2], Anna Triantafillou[3], Aggeliki Androutsopoulou[1] & Euripidis Loukis[1]

[1] Department of Information and Communication Systems Engineering
University of Aegean
Gorgyras and Palama Str.
83200 Karlovassi, Samos, Greece
[2] National Center for Scientific Research (NCSR) "Demokritos"
P. Grigoriou and Neapoleos Str.
15310 Ag. Paraskevi Attikis, Athens, Greece
[3] Athens Technology Center
Rizariou Str. 10
15233 Halandri, Athens, Greece

yannis@aegean.gr
vangelis@iit.demokritos.gr
a.triantafillou@atc.gr
ag.andr@aegean.gr
eloukis@aegean.gr

Abstract: While the first generation of e-participation has been based on official e-participation spaces owned and operated by government, the second one is oriented towards exploiting the highly popular web 2.0 social media for performing 'crowd-sourcing' of policy-related knowledge, opinions and ideas from citizens, through posting relevant policy-related content to some social media and then retrieving and processing citizens' interactions with it. Recently, the idea of a third generation of e-participation has been proposed, which is based on a more 'passive' form of crowd-sourcing in social media, through automated passive search by government agencies for content on a public policy under discussion, that has been created in a large number of predefined relevant web 2.0 sources (e.g. political blogs, news websites, facebookand twitter accounts) by citizens freely, without any direct stimulation by government, retrieval and sophisticated processing of this content. In this paper we analyze and elaborate this idea, based on cooperation with potential users experienced in the design of public policies, through a combination of quantitative and qualitative techniques. Initially, the main roles required for the practical application of this concept are identified, and then the functional requirements of each of them are determined. Finally, based on these functional requirements the architecture of a central platform supporting the application of this concept is designed.

1 Introduction

In the last decade there has been a lot of effort and investment by governments in order to exploit the capabilities of information and communication technologies (ICT) for establishing new electronic channels of communication with the citizens, and enable a more open, citizen-centric and participatory public policy making model [LMC11]. The first generation of e-participation has been based on official e-participation spaces owned and operated by government agencies; this first e-participation paradigm was highly government-controlled, as these official e-participation spaces defined and controlled totally the forms, the rules and the topics of electronic discussions taking place there, and has not been extensively used by citizens [Ch09], [FM10]. The emergence of the web 2.0 social media and their success in attracting big numbers of users lead to the gradual emergence of a second generation of e-participation, whichis oriented towards the exploitation of the social media for performing crowd-sourcing of policy-related knowledge, opinions and ideas from citizens, through posting relevant content to some social media and then retrieving and processing citizens' interactions with it [CL12]. In this second e-participation paradigm governments continue defining and controlling the discussion topics to some extent, through making postings and initiating discussions in various social media on the topics they choose, but do not control the forms and the rules of the discussion, which are the ones defined by the social media.

Recently, the idea of a third generation of e-participation has been proposed, which is based on a more 'passive' form of crowd-sourcing: it includesautomated passive search by government agencies for content on a public policy, which has been created in a large number of predefined relevant web 2.0 sources (e.g. political blogs, news websites, facebook and twitter accounts) by citizens freely, without any direct stimulation by government, and then retrieval and sophisticated processing of this content[Ch12]. This new e-participation paradigm is characterized by even less government control and more citizens' control: governments control neither the discussion topic, nor the discussion forms and rules.

In this paper we analyze and elaborate this novelidea, based on cooperation with potential users experienced in the design of public policies, througha combination of quantitative and qualitative techniques. Initially, the main roles which are necessary for the practical application of this concept are identified, and then the functional requirements of each of them are determined. Finally, based on these requirements the architecture of a central platform supporting the application of this concept is designed.The research presented in this paper has been conducted as part of the research project NOMAD ('Policy Formulation and Validation through non moderated crowd-sourcing', which is partially funded by the 'ICT for Governance and Policy Modelling'-see http://www.nomad-project.eu/) partially financed by the European Commission.

The paper is structured in six sections. The following section 2 outlines the background of this concept, whereas in section 3 is described the methodology we followed in order to identify the main actors and the corresponding functional requirements. Then, in section 4 the functional requirements are outlined, and section 5 presents the

technological architecture of an ICT platform that fulfills these requirements and supports the practical application of this concept. Finally, section 6 summarizes the conclusions and proposes future research directions.

2 Background

Management literature has been discussing for long time the capability of a large network of peopleconnected throughICT, termed as 'crowd', to perform successfully difficult design and problem solving activities ([Le97],[ML04]).This collective intelligence has recently started being exploited systematically by organizations. This practice is referred to as 'crowd-sourcing', defined as "the act of a company or institution taking a function once performed by employees andoutsourcing it to an undefined (and generally large) network of people in the form of an open call. This can take the form of peer-production (when the job is performed collaboratively), but is also often undertaken by sole individuals" [Ho06].

Initially,crowd-sourcing was introduced in some creative and industry sectors as an innovative practice for new products' design and problem solving by large networks of people instead of individual professionals, exploiting the distribution and diversity of intelligence ingroups or crowds, known as "crowd wisdom"([Le97],[Su04]). Eventually, more and more firmsfrom various sectors started addressing complex tasksthrough crowd-sourcing solutions, whichoften outperformed in-house professional solutions ([Su04], [Ho06], [Br08], [Br12]). Research in this area has lead to the identification of several different typologies of crowd-sourcing practices [Br12]; however, most of them share some common characteristics: they are based on an'open call' by an initiating firm to all interested individuals for information, knowledge, ideas and solutions for a particular problem, followed by submissions by numerous individuals, which are evaluated by the firm, and finally the best ones are given some predefined reward.

These crowd-sourcing ideas were initially developed and applied in the private sector, but recently started being applied in the public sector as well. This gave rise to the so-called 'citizen-sourcing', which is defined asthe process of gathering citizens' knowledge, ideas, opinions and needs in order to address various types ofsocietal problemsthatgovernment agencies face ([Bo07], [LT08],[HI10],[Na12]). Most of these first attempts to apply crowd-sourcing ideas in government took the form of 'active crowd-sourcing', in which a government agency makes one or more postings in their own websites or in their social media accounts with a question or a problem, and solicits relevant information, knowledge, ideas, suggestions or opinions from citizens.

The present paper analyses and elaborates the idea of 'passive crowd-sourcing'bygovernment agencies as a means for improving their public policy making processes. It constitutes an attempt to adapt and apply in the context of government policy making processes the idea of 'digital reputation management', which was originally created and applied in the private sector marketing domain ([YS02], [Ho05], [RMF11]).

3 Methodology

In order to identify the main roles required for the practical application of this passive crowdsouring concept in the policy making processes of government agencies, and determine the functional requirements of each from asupporting ICT platform,we cooperated with potential users experienced in the design of public policies in the countries of the three user partners of the NOMAD project (Austria, Greece, UK). For this purpose we used a combination of quantitative and qualitative techniques:

1. Initially four application scenarios of this concept were developed by the user partners of the project. Each scenario constitutes a detailed realistic example of how this concept might be applied for supporting the formulation of a particular public policy.

2. A questionnaire was then distributed electronically to a sample population of potentialusers, aimingwas to determine current policy making processes and work practices, relevant needs and also opinions about thisconcept and the above application scenarios.

3. Furthermore focus groups and workshopswere organized with the participation of potential users, which included in-depth discussionsabout this new concept, ways of its practical implementation, required functionalities and also possible problems and barriers to its adoption by government agencies.

4. Finally, in-depth semi-structured interviews were organized this new concept, focusing on its applicability, advantages, disadvantages and ideas for improving it.

In the above activities participated a mix of different types of actors involved in public policy making: decision makers, policy advisors, public servants, representatives of non-governmental organisations and trade unions. This cooperative approach allowed us to elaborate this idea of a passive crowdsourcing for supporting public policy making, to specify realistically how it might practically work, to collect attitudes, needs and requirements from a wide range of potential users, and to understand their expectations.

4 Requirements

Using the above methodology we identified four main roles required for the practical application of this concept, and the functional requirements of each of them:

I. The “Domain Expert” is a role who will be in charge of creating a domain model, which will consist of the main keywords/terms (key concepts) of the specific domain, for which a policy is intended, and the relations among them.An example of such a domain model for the energy domain developed as part of our work for the "Greek Strategy for Energy Planning" is shown below in Figure 1.

Figure 1: Domain Model for the "Greek Strategy for Energy Planning"

Therefore the domain expert will be able to create a domain model (by defining the main keywords/terms of the domain and the relations between them), store it, convert it to an ontology OWL file, and alsoaccess and use or modify existingdomain models. These terms of the domain model will be used for searching for relevant content created in a large number of predefined relevant web 2.0 sources (mainly political blogs, news websites and facebook and twitter accounts of them).

II. The "Policy Advisor" is a role who will be in charge of creating a policy model, which will be based on an existing model of the corresponding domain, adding to its nodes 'policy statements' (meant as specific policy objectives and actions/interventions that a policy includes) and arguments in favour or against them. Therefore the policy advisor will be able to create a policy model (by adding policy statements, both policy objectives and actions/interventions, to the nodes of a domain model, and also arguments on them), and also access and use or modify existingpolicy models. Furthermore, he/she will be able to have a basic view of the results of content search in the above predefined sources with respect to a specific domain or policy model, including the most frequently mentioned topics in the searched content sources concerning this domain or policy model, visualised as a tag cloud; also, he/she will be able to view the tag cloud that corresponds to a particular time period, content sources subset and audience (e.g. focusing on young citizens only).

III. The "Policy Maker" is a central role who will be provided an extensive set of capabilities for advanced searches regarding adomain or a particular policy. He/she will be able to view not only the most frequently mentioned topicswith respect to the domain or policy model in the above sources, visualised as a tag cloud, but also the public sentiments (positive or negative) towards them,and the differentiations of both topics and sentiments over time and across different citizens'groups, as well as their forecasted future evolution. Furthermore, he/she will be able to view the public sentiments towards the main policy statements (i.e. policy objectives and measures/interventions),and the arguments on them, which are included in a policy model, and also their differentiation over time and across different citizens'groups, as well as their forecasted future evolution. Finally, he/she will be able to compare the public sentiment towards entities

of different policy models in the same domain in order to observe the impact of alterations in the policy under discussion.

IV. The "Technical Administrator" is a role responsible for setting up and monitoring the ICT platform from a technical viewpoint, for ensuring its proper operation, for solving technological problems and for managing users and roles (rights, permissions, accounts).

5 Technological Architecture

Based on the above functional requirements the technological architecture of an ICT platform that fulfils them, and in general supports the application of this passive crowdsourcing concept in government, was designed. The objective of the technical design was to provide this functionality with an acceptable response time. Since this could not be achieved through online retrieval of content from the multiple and heterogeneous predefined sources (multiple political blogs, news websites, facebook and twitter accounts) and processing of it performed at the time a user initiates a search, the only solution was to perform a scan of the predefined sources at given intervals (e.g. every 6 hours) in order to retrieve new content, to store it in repositories and then process it. Whenever the user performs a search, the results will be produced in a very short time, using these repositories. This separation between sources scanning and content processing on one hand and users' searches processing on the other, allows a low response time and at the same time sufficiently 'fresh' content for policy makers.

The above design leads to a three layers' technological architecture of the platform,which consists of a storage layer, a processing layer and a presentation layer (see Figure 2). Each of them includes a number of components performing various different tasks, which act as services coordinated by an orchestration component.

In particular, the Data Storage Layer includes the repositories where the raw and processed content is stored:

- The *Content Repository:*it stores the raw contentretrieved from the web 2.0 sources, the cleaned content derived from the raw data, the content uploaded by users and the results of the linguistic analysis associated with each content unit.
- The *Model Repository:*it stores in a structured form of the domain and policy models entered by users with domain expert and policy advisor roles.
- The *Metadata Repository:* it stores the metadata retrieved or calculated for the content our sources.
- The *Thematic Catalogues:*it stores a representation of the thematic categories used by the platform in order to characterise each content unit.
- The *Users Repository:*it contains information about the roles andthe users of the platform.

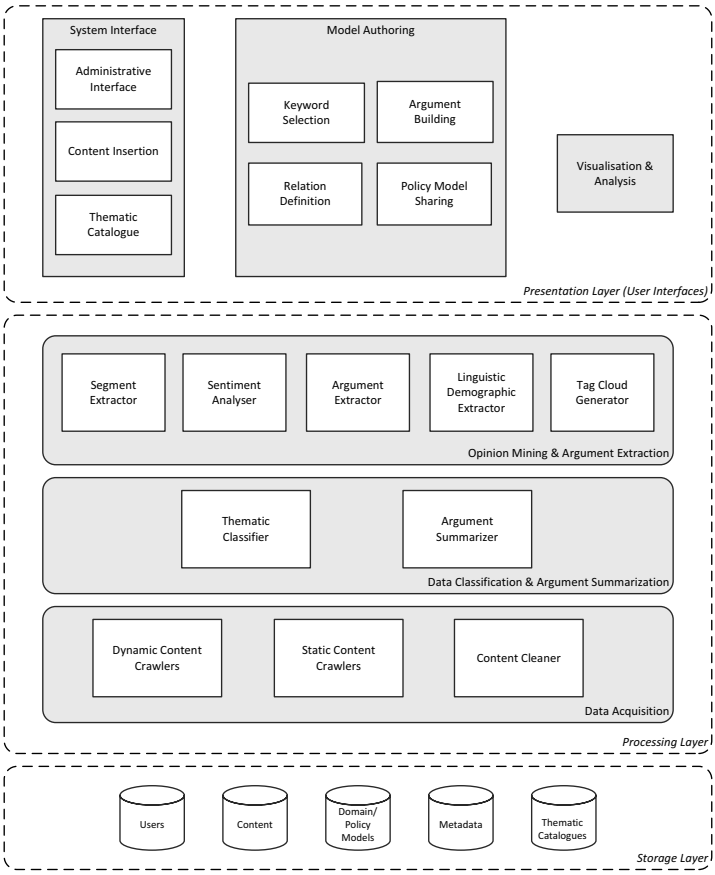

Figure 2: Overview of the technological architecture of the platform

The Processing Layer includes all the components that retrieve and process the content from the predefined sources, which are organized in three sub-layers:

• The *Data Acquisition Layer*, which includes the crawling components for fetching content from the sources, as well as the modules responsible for cleaning the fetched content and obtaining the actual textual information from it (Static Content Crawlers, Dynamic Content Crawlers and Content Cleaner).

• The *Data Classification & Argument Summarization Layer*, which includes (a) the Thematic Classifier, which processes the available content and associates it with one ore more of the defined thematic categories, and (b) the Result Summarizer, which processes the available results and provides a summarization that allows their presentation in a condensed manner.

• The *Argument Extraction & Opinion Mining Layer*, which includes all the components that process the available content and extract segments, arguments and sentiments (Segment Extractor, Argument Extractor, Sentiment Analyzer, Linguistic Demographic Extractor, Tag Cloud Generator).

The Presentation Layer includes all the components that require input from the user or present to him/her the results:

- The *Thematic Catalogue* isan interface for entering or updating the available thematic categories and also terms associated with each category.
- The *Keyword Selection* interface, which allows entering keywords/terms for creating domain models
- The *Relation Definition* interface, which allows the user to introduce relations between the above keywords/termsfor the definition of domain models
- The *Argument Building* interface, which offers a visualisation of a policy model and allows the user to insert in natural language policy statements (policy objectives and policy measures/interventions) and arguments supporting or objecting to them.
- The *Policy Model Sharing* interface, which provides a catalogue of the policy models created by the user and allows defining them as visible to others.
- The *Admin interface*, which provides the means to an administrator to manage the configurable aspects of the system.
- The *Visualisation and Analysis module*, which utilizes the results of the processing layer in order to provide the user with a view of domain and policy models, and also various visualizations of the results of users' searches, enabling also the latter to enter corresponding sources, demographic characteristics and time periods.

6 Conclusions

Recent literature has proposed the use of 'passive crowdsourcing' by government agencies in their public policy processes, based on the search for content on a public policy that has been created by citizens freely, without any direct stimulation by government, in a large number of predefined relevant web 2.0 sources (e.g. political blogs, news websites, facebook and twitter accounts). In the previous sections of this paper we analyze and elaborate this idea in cooperation with potential usersexperienced in the design of public policies, using several different quantitative and qualitative techniques.

It has been concluded that the practical application of this new concept will require four main roles - domain expert, policy advisor, policy maker and technical administration – and for each of them the functional requirements from a central supporting ICT platform were determined. The inherent complexity and multidimensionality ofpublic policies makes it difficult to search for relevant content in the predefined web 2.0 sources, and this difficulty can be addressed through the creationof a model with the main terms of the domaineach policy is intended for, and the relations among them; on the appropriate nodes of this domain model can be added the main policy statements (policy objectives and policy interventions) and posistive/negative arguments on them, and this leads to a model of the policy, which can be used for searching for relevant content in the predefined web 2.0 sources. Also, a sequence of processing steps of the retrieved content has been designed, which results in the identification of the main topics mentioned by citizens in the above sources, the corresponding sentiments (positive or negative), their differentiations over time and across different citizens'groups, and also a forecast of

their future evolution. The above provide a ‘business model’ for the application of this passive crowdsourcing idea in government, which describes how it canpracticallywork and what kind of value it can offer to policy makers.

Finally, based on the above requirements the technological architecture of an ICT platform that fulfils them, and in general supports the application of this passive crowdsourcing concept in government, was designed. In order to achieve a low response time and at the same time sufficiently ‘fresh’ content for policy makers, its design should be characterised byseparation between sources scanning and content processing on one hand and users’ searches processing on the other.This leads to a three layers’ technological architecture of the platform, which consists of a storage layer, a processing layer and a presentation layer. Each layer includes a number of components, with each of them performing a different task, which act as services coordinated by an orchestration component.

Further research is required for evaluating the above ‘business model’ and functional and technological architecture, based on real-life pilot applications of it in various policy domains, which will probably result in modifications and improvements. This is already in progress as part of the abovementioned project NOMAD.

References

[Bo07] Bovaird, T.: Beyond engagement and participation: User and community coproduction of public services. Public Administration Review 67.5, 2007; pp. 846-860.

[Br08] Brabham, D. C.: Crowdsourcing as a Model for Problem Solving - An Introduction and Cases. Convergence: The International Journal of Research into New Media Technologies 14.1, 2008; pp.75-90.

[Br12] Brabham, D. C.: Crowdsourcing: A Model for Leveraging Online Communities. In: (Delwiche, A. and J. Henderson): The Routledge Handbook of Participative Cultures. Routledge, 2012.

[Ch09] Chadwick, A.: Web 2.0: New Challenges for the Study of E-Democracy in an Era of Informational Exuberance. I/S: A Journal of Law and Policy for the Information Society, 5 (1), 2009; pp.9-41.

[Ch12] Charalabidis, Y. et. al.: Public Policy Formulation Through Non-Moderated Crowdsourcing in Social Media. IFIP Fourth International Conference on e-Participation - ePart 2012, 3-6 September, 2012, Kristiansand, Norway, 2012.

[CL12] Charalabidis, Y.; Loukis, E.: Participative Public Policy Making Through Multiple Social Media Platforms Utilization.International Journal of Electronic Government Research. SCOPUS, 8(3), 2012; pp. 78-97.

[FM10] Ferro, E.; Molinari, F.: Making Sense of Gov 2.0 Strategies: No Citizens, No Party. Journal of eDemocracy and Open Government, 2 (1), 2010; pp. 56-68.

[HI10] Hilgers, D.; Ihl C.: Citizensourcing:Applying the concept of open innovation to the public sector. The International Journal of Public Participation 4.1, 2010; pp. 67-88.

[Ho05] Howard, C.: The Policy Cycle: A Model of Post-Machiavellian Policy Making. Australian 64, 2005; pp. 3-13.

[Ho06] Howe, J.: The Rise of Crowdsourcing. Wired 4.6, 2006.

[Le97] Lévy, P.: Collective Intelligence: Mankind's Emerging World in Cyberspace, Plenum, New York, 1997.

[LMC11] Loukis, E.; Macintosh, A.; Charalabidis, Y.: Editorial of the Special Issue on E-Participation in Southern Europe and the Balkans: Issues of democracy and participation via electronic media. Journal of Balkan and Near East Studies, 13 (1), 2011; pp. 1-12.

[LT08] Lukensmeyer, C. J.; Torres L. H.; Citizensourcing: Citizen participation in a networked nation. Civic Engagement in a Network Society, 2008;pp. 207-233.

[ML04] Mau, B.; Leonard, J.; The Institute Without Boundaries: Massive Change. Phaidon, New York, 2004.

[Na12] Nam, T.; Suggesting frameworks of citizen-sourcing via Government 2.0. Government Information Quarterly 29, 2012; pp.19-20.

[RMF11] Raguseo, E.; Mosconi P.; Ferro E.: Reputation management as a lever of public sector innovation. Conference for E-Democracy & Open Government (CEDEM 2011).Krems, Austria, 2011.

[Su04] Surowiecki, J.;The Wisdom of Crowds: Why the Many are Smarter than the Few and How Collective Wisdom Shapes Business, Economies, Societies, and Nations. Doubleday, New York, 2004.

[YS02] Yu, B.; Singh M.; Distributed Reputation Management for Electronic Commerce.Computational Intelligence 18.4 2002; pp. 535-549.