An e-Participation Support System for Regional Communities Based on Linked Open Data, Classification and Clustering

Robin M. E. Swezey, Hiroyuki Sano, Norifumi Hirata, Shun Shiramatsu, Tadachika Ozono and Toramatsu Shintani
Dept. of Computer Science and Engineering,
Graduate School of Enginnering,
Nagoya Institute of Technology
Aichi, Japan 466–8555
Email: {robin, hsano, nori, siramatu, ozono, tora}@toralab.ics.nitech.ac.jp

***Abstract*—In this paper, we present an e-Participation Web system for various platforms and mobile phones, based on linked open data, classification and clustering. The system aims at supporting citizen e-Participation in ongoing regional debates by gathering and openly publishing news and opinions from the Web for easy comprehension and commenting. Our study helps us define relevant evaluation criteria for an adequate citizen discussion system in the new context of open government, the Web, and mobile computing. We present the system, O2, and its application citispe@k, as well as its underlying components: ontology structure, classification and clustering. We then conduct a comparison with existing systems and find that our system is a better approach for efficient citizen e-participation when compared to current existing systems.**

I. Introduction

We developed an e-Participation Web platform, O_2, for regional communities. The platform aims at supporting citizen e-Participation in ongoing regional debates by gathering and openly publishing news and opinions. Structuring citizens' awareness of the regional issues and sharing the structured data are two requirements to conduct productive discussions about various issues. O_2 consists of three tools: Sophia, SOCIA, and citispe@k. Sophia is a mining and intelligent pre-processing platform which classifies and clusters news articles and tweets. SOCIA is a data set and its ontology for debate, based on Linked Open Data (LOD), to archive information and discussion about events occurring in regional communities. Citispe@k is an application to support the discussion of regional issues identified by Sophia using SOCIA.

In order to gain better engagement and involvement from citizens, information from the Web (e.g articles, blogs, tweets) needs to be thoroughly classified by region, and then presented to citizens in an understandable way. Using our platform and ontology, news and opinions are structured and linked with regional issues, and the data is openly published on the Web using the LOD-based ontology of SOCIA. Through this process, e-Participative data becomes re-usable and transparent. Transparency is a requirement of Government 2.0 initiatives.

Data mined from the Web is structured in the form of events by region, which are then used as discussion seeds to further build SOCIA. Citizens then create discussion topics out of each seed, e.g a cluster of news related to the same event, and input their opinions by using the system, among other functionality.

The system first collects news articles and microblog posts (in this work, tweets) along with necessary metadata (dates, emission sources, etc). It then classifies this crawled data by region and filters out noise irrelevant to the interest of regional communities or current events. Next, the system extracts target events from the news articles and microblogs, and links them using the ontology. Citizens can then add further links to events, news articles and microblogs, by creating relevant topics and debate about them by inputting their opinions, polling, or sharing further resources. Those resources and new links are also incorporated in the data set, as are the opinions and the discussion. This creates a virtuous circle where the intelligent platform, by creating understandable and relevant discussion seeds, involves citizens in e-Participation. The citizens add further data to the data set, making it grow over time, and this data can be used as input again (e.g for training better learning models or developing better ontologies).

The rest of this paper is organized as follows. In section II, we briefly describe related works. The details of the data set SOCIA are introduced in section III. The application of SOCIA is described in section IV. We compare our system and current systems which can be used for debates in section V to insist superiority of our system. And in section VI, we summarize our contributions and conclude the paper.

II. Related Works

In this section, we present related works to introduce the context in which we will present and evaluate our system for regional citizen participation involvement.

A. e-Government and e-Participation

e-Government, as defined in [1], consists in the employment of the Internet and the world-wide-web for delivering govern-

ment information and services to the citizens. Mainly, it refers to the use of new technologies by governments to reach and interact with citizens.

Actual use of these tools to extend the scope of e-Government by including citizen engagement and participation in governance, i.e use of information and communication technologies to achieve better governance, is referred to as e-Governance. Finally, use of e-Participation tools in the decision-making process of democratic government organizations is referred to as e-Democracy [2].

e-Participation is the use of information and communication technologies to broaden and deepen political participation by enabling citizens to connect with one another and with their elected representative [2].

B. e-Participation Tools and Technologies

Wimmer et al. [3], with Macintosh et al. [4], group the existing e-Participation tools into three main clusters: core e-participation tools, tools used in e-participation but not specific to e-participation, and basic tools to support e-participation.

Core e-participation tools are actually tools that use a goal-specific definition relatively to e-government: e-Participation chatrooms, e-Participation message boards, decision-making games, virtual communities, online surgeries, e-Panels, e-Petitioning, e-Deliberative Polling, e-Consultation, e-Voting, suggestion tools for planning procedures. The first four are not specific to e-Participation but in a goal-oriented domain they could be regrouped as e-Participation discussion and involvement tools, which is what O_2 is. As well, tools used in e-Participation are technology-specific tools, namely webcasts, podcasts, wikis, blogs, quick polls, surveys, GIS-tools. Tools of support are actually legacy technology of the environment of e-Participation, used by citizens on their own account but not provided by the system: search engines, alert services, newsletters, FAQs, portals, groupware tools.

This study helps us give a more precise context to O_2, which is an e-Participation discussion and involvement tool (goal) with an innovative technological approach relying on structuring, modeling and presentation of data (technology) with support by using outside data. In the later comparison with other tools that we conduct in Section V, this helps us define relevant comparison criteria for O_2.

C. Open Meeting and Further Initiatives

Open Meeting constitutes a special case of an e-Participation support system. Developed for vice president Al Gore's Open Meeting project on National Performance Review, the Open Meeting system [5] made use of knowledge representation, hypertext grammar, and rules for commenting. The system was meant to empower users with the ability to conduct policy conversations without any agency boundaries. The research conducted for the development of the system first identified the interactions needed for productive discussion in a large group. For example, specialists could only access texts relevant from their interests and link comments on those texts. Open Meeting was built around a very precise structure of debate, with a hypertext grammar that made it possible, whereas systems such as newsgroups and traditional bulletin boards were not suited for such participative debate. In the context of regional productive citizen discussion, we thus consider Open Meeting as an ground-laying and inspiring work for SOCIA, since it shows that relevant discussion structuring through technological innovation leads to effective citizen collaboration and e-Participation.

Next in line is the MIT Deliberatorium [6]. Deliberatorium is a system to enable better collaborative deliberation. It works through systematic exploration evaluation and convergence on solution ideas, including stakeholders and experts. Different from systems such as message boards where interaction is time-centric, it aims at solving the problems of scattering of points on a topic, balkanism (the clustering of like-minded users on threads), and the soapbox problem where the last to speak is the last to be heard, and small voices tend to get left out. To address this issue of noise, Deliberatorium relies on argument mapping, in a manner very similar to Open Meeting. This results in no scattering, no soapbox problem, and bias towards well-founded arguments. While Deliberatorium structures discussion over long-time spanning topics, O_2 chooses a more event-centric approach. The reason is that events tend to create a flock of individuals, described as swarming [7], as a reaction to punctual events that directly concerns them. When repeated, ad-hoc movements born from swarming can engage in participation on a more long-term basis. Past research was also conducted on the use of swarming towards service issues in hope to create a participation habit [8]. Since O_2 also aims at structuring debate, it adds structure to the discussion through the use of various tags (see Sec IV).

Finally, another work of interest is Cohere [9]. Cohere is a social, semantic web application described as a working prototype based on the rationale of contested collective intelligence. Cohere focuses on sensemaking, and its is to help users make sense out of data, by connecting ideas and annotating Web pages, then linking them through an ontology. According to its own definition, it sits at the intersection of Web annotation, social bookmarking, and mindmapping. Cohere differs from our project in the sense that it is strongly based on idea connection and knowledge mapping and sharing, whereas O_2 focuses on regional citizen participation using current events as discussion seeds.

III. Platform:O_2/SOCIA

A. O_2/Sophia/SOCIA

O_2 is a Web platform for citizen participation in debates about regional issues. O_2 is an abbreviation for Open Opinion. Fig. 1 shows the outline of O_2 platform. O_2 has three stages. In stage (1), the mining and pre-processing system Sophia crawls the Web and gathers informations such as news articles or microblogs, which can be used for debates from the Web. In stage (2), the system tries geographical classification and event clustering to structure the gathered data. Relevant data of interest is then structured and stored in the data set SOCIA according to the SOCIA ontology, as openly published Linked

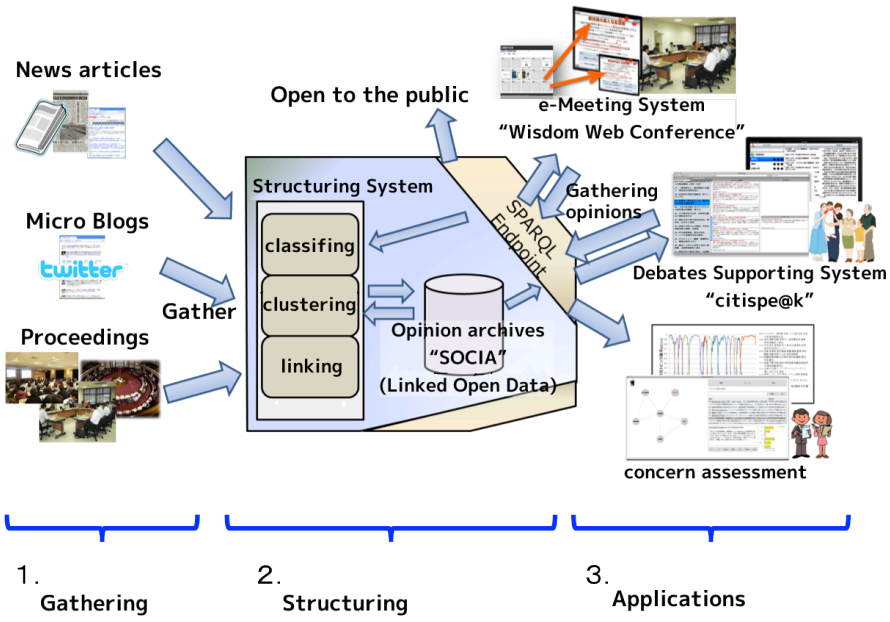

Fig. 1. Outline of O_2

Open Data. Stage (3) is the application of SOCIA's purpose, debate support. Several applications can actually branch to Sophia/SOCIA, not only debate supporting systems but also an e-Meeting system we developed. In this paper, we focus on the e-Participation system for supporting debates about regional issues.

B. Classification by Region

After mining, we perform classification of news articles and tweets by geography (against the 47 prefectures of Japan). To this end, we use Naive Bayes text classification. The metric chosen is the Tf-Idf metric as follows:

$$\mathrm{tfidf}(t, \mathbf{d}, D) = \mathrm{tf}(t, \mathbf{d}) \times \mathrm{idf}(t, D) \tag{1}$$

with

$$\mathrm{tf}(t, \mathbf{d}) = \mathbf{d} \cdot \mathbf{t}$$
$$\mathrm{idf}(t, D) = \log \frac{|D|}{1 + |\{\mathbf{d} \in D : \mathbf{d} \cdot \mathbf{t} > 0\}|}$$

where $|D|$ is the cardinality of the corpus of documents, and $|\{\mathbf{d} \in D : \mathbf{d} \cdot \mathbf{t} > 0\}|$ is the number of documents $\mathbf{d}$ where the term t appears, with $\mathbf{t}$ its unit vector representation.

Naive Bayes itself assumes that words are drawn independently from a multinomial distribution and that, given a class label c from the prefectures of Japan:

$$\mathrm{P}(c|\mathbf{d}) = \frac{P(\mathbf{d}|c) \times P(c)}{P(\mathbf{d})} \tag{2}$$

However, straightforward use of Tf-Idf metrics with Naive Bayes not being efficient and accurate enough, we use bi-gram classification where the $\mathbf{t}$'s are bi-morheme combination

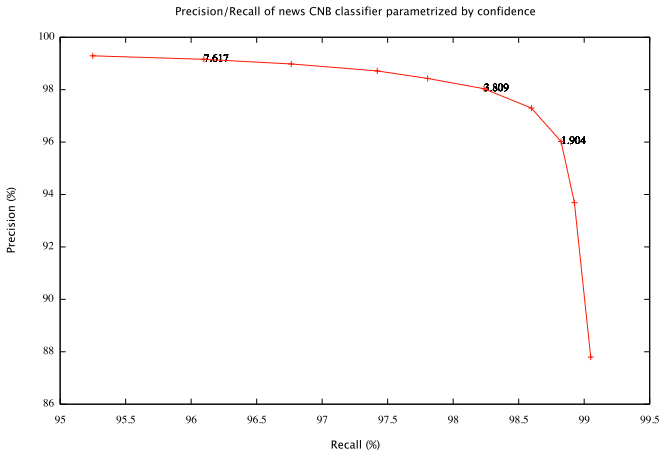

Fig. 2. Precision/Recall parametrized by confidence in classification of news articles

features drawn from the text. The Naive Bayes algorithm as well does not perform well enough for text classification, so we instead use a Transformed Weight-normalized Complementary Naive Bayes algorithm [10]. Further improvements of the classification algorithm are the subject of another work. The score for classifying a document d into a class c is calculated as:

$$\hat{\mathrm{w}}(c|\mathbf{d}) = \log P(c|\mathbf{d}) - \sum_t \mathrm{tf}(t, \mathbf{d}) log(\frac{1 + \sum_{k=1}^{|C|} \mathrm{tf}(t, c_k)}{N + \sum_{k=1}^{|C|} \sum_{x=1}^{N} \mathrm{tf}(x, c_k)}) \tag{3}$$

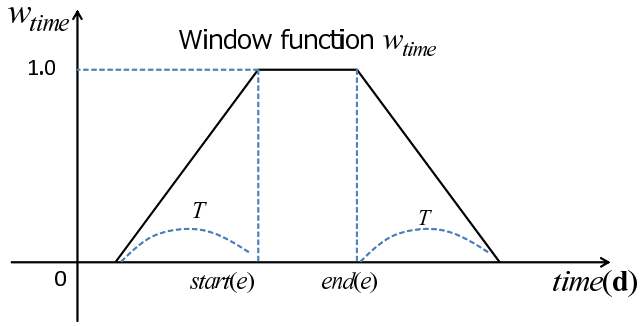

Fig. 3. Window function for considering dates/times the news articles were published

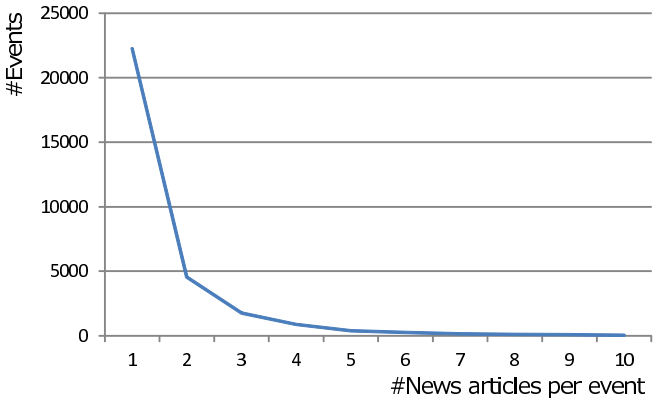

Fig. 4. Distribution of news article counts per event

To decide whether or not contents should be filtered out, we use a confidence threshold where the classifier confidence is determined by:

$$\gamma(\mathbf{d}) = \hat{\mathrm{w}}(c_1|\mathbf{d}) - \hat{\mathrm{w}}(c_2|\mathbf{d}) \tag{4}$$

with

$$\mathrm{c}_1 = \arg\max_{c \in C} \hat{\mathrm{w}}(c|\mathbf{d}), \mathrm{c}_2 = \arg\max_{c \in C \setminus \{c_1\}} \hat{\mathrm{w}}(c|\mathbf{d})$$

where c_1 and c_2 are respectively the first and second class labels where $\mathbf{d}$ weighs the most, the first two classes for which the classifier is most confident.

We conducted a classification experiment through varying threshold of confidence value, using 8,811 news articles related to Japanese prefectures crawled from Yahoo! Japan News[1] during Jun. 13 to Jul. 12, 2011.

Figure 2 shows the confidence of the classifier when tested with random noise text that is likely to be mined in production. We can then decide an appropriate threshold by favoring precision over recall. This strategy enables us to filter contents irrelevant to the interests of regional communities.

C. Clustering by Events

To extract events from news articles, the system uses a cosine measure based on tf·idf. Each dimension of a document vector corresponds to a separate term, and each component corresponds to an evaluation of the term. However, the system performs the calculation by assigning a certain weight to the term based on the time the articles are delivered.

The similarities $\mathrm{sim}(d, e)$ with a new news article d and each event e on SOCIA are calculated when the article is saved on SOCIA. If the similarity is greater than a threshold θ defined empirically, the article is linked to the event. If all similarities are less than the predefined threshold, the system makes a new event from the article. The similarity is calculated with bag-of-words vectors of d and e consisting of TF*IDF values, term weights w_{term}, and window function w_{time} (shown in Fig. 3) as follows:

$$\mathrm{sim}(\mathbf{d}, e) = w_{\mathrm{time}}(time(\mathbf{d})|e) \times \frac{\sum_t w_{\mathrm{term}}(t|\mathbf{d})\mathrm{tfidf}(t, \mathbf{d}, D)\mathrm{tfidf}(t, e, D)}{\sqrt{\sum_t w_{\mathrm{term}}(t|\mathbf{d})^2 \mathrm{tfidf}(t, \mathbf{d}, D)^2}\sqrt{\sum_t \mathrm{tfidf}(t, e, D)^2}} \tag{5}$$

with

$$w_{\mathrm{term}}(t|\mathbf{d}) = \begin{cases} \alpha & > 1, \text{ if the term } t \text{ appears in } \mathbf{d} \text{ 's title} \\ 1 & \text{otherwise,} \end{cases}$$

$$w_{\mathrm{time}}(time(\mathbf{d})|e) = \begin{cases} 1 & \text{if } start(e) < time(\mathbf{d}) < end(e) \\ \frac{end(e)+T-time(\mathbf{d})}{T} & \text{if } end(e) < time(\mathbf{d}) < end(e) + T \\ \frac{time(\mathbf{d})-(start(e)-T)}{T} & \text{if } start(e) - T < time(\mathbf{d}) \\ & < start(e) \\ 0 & \text{otherwise,} \end{cases}$$

where let $time(\mathbf{d})$ be a published time of $\mathbf{d}$, $start(e)$ be a published time of the earliest article included in e, and $end(e)$ be a published time of the latest article included in e. The similarity threshold and the weight of terms appear in news title were empirically set as follows: $\theta = 0.4$ and $\alpha = 3.0$.

SOCIA stored 54,854 news articles and about 13,000 ones classified to prefictures[2]. Fig. 4, the distribution of news articles per event, shows that 34,971 events are extracted through clustering the 54,854 articles.

The system also calcurates the similarity scores between all events stored on SOCIA. For example, the similarity between event e_i and event e_j gets greater score than a threshold, the system treats that event e_i is related to event e_j. The similarity is formulated as follows:

$$sim(e_i, e_j) = \frac{\sum_{k=1}^{N} w_1(e_i, n_k) \cdot \sum_{k=1}^{N} w_2(n_k, e_j)}{\sqrt{\sum_{k=1}^{N} w_1(e_i, n_k)^2} \cdot \sqrt{\sum_{k=1}^{N} w_2(n_k, e_j)^2}} \tag{6}$$

[1]http://headlines.yahoo.co.jp/hl?c=loc

[2]The number of news articles stored in SOCIA was counted on Mar. 16, 2012. It has been constantly increasing.

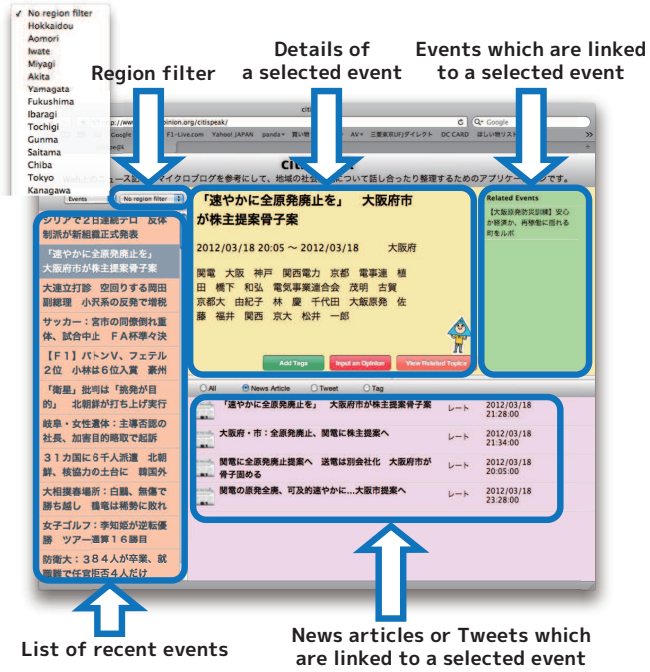

Fig. 5. The screenshot of citispe@k

Fig. 6. Add an opinion to the news article which a user is viewing

In this formula, $n_k(k = 1, 2, \cdots N)$ are news articles linked to both events e_i and e_j. $w_1(e_i, n_k)$ is the similarity of e_i and n_k calucrated by cosine measure, $w_2(n_k, e_j)$ is that of e_j and n_k.

IV. Application:citispe@k

A. Summary of the system

Citispe@k is a debate support system based on SOCIA, implemented as a Web application, usable on Web browsers. For mobility and reach, Web browsers running on smart phones and tablets are supported.

The origin of the word citispe@k is that citizens speak about social issues and current events of the regions they live in. Users can discuss about or sort out regional issues with referencing news articles, tweets or other relevant resources on the Web by using citispe@k. By creating discussion topics or inputting opinions on the system, those topics or opinions are also stored as Linked Open Data in SOCIA, adding more to Linked Open Data naturally.

Fig.5 shows a screenshot of citispe@k. The screenshot has lists of events or related information. Events recently updated are listed on the left side of the screenshot. First the system shows all events, but users can limit the list to show only events related to their region. When users select an event from the list, information about the event is shown on the right side of the screenshot. Information consists of news articles, tweets, events related to the event. Those resources can be easily shown and visualized in an iFrame without leaving the system. Fig.6 shows a screenshot of a user selecting a news article from the list. In Fig.6, the system header has three buttons. The button "Add Tags" is used to add tags to the news article, the button "Input an Opinion" is used to add a comment to the news article and the button "View related topics" is used to see topics which are linked to the news article. The added comment is treated and structured as an opinion. The comment can also be posted to Twitter (via @citispeak for now) to further its reach, and is stored on SOCIA. Optional Twitter accounts will be supported soon.

Users can create discussion topics with relations to events, news articles and tweets. "View related topics" button in Fig.7 is used to see topics which are related to the event when users are viewing a event. Users can make a new discussion topic about the event by clicking the "Make a new topic" button. The cycle of the discussions in citispe@k is that users browse events, get related topics about an event, and add their opinion to the topic they are interested in. The system supports to add some Web contents to topics as information sources for discussion, not only add opinions to topics.

B. Practical use for concern assessment

We define concern assessment as follows: to analyze a trend in citizens' awareness or anxieties of the social issues. For example, committee for science verification about road construction in Aioiyama-Ryokuchi Park in Nagoya City, analyzes road construction[3]. A report of construction analysis was made based on five criterias, "economic chance", "life, educational or cultural chance", "safe, security" and so on. Thus, classifing opinions based on criterias is effective for concern adjustment.

Citispe@k provides above-mentioned tags. Concretely, users can add tags composed with criteria and polarity to them such as "Environment +" or "Environment -". citispe@k also provides tags which can be used to express the kind of opinions like "Question", "Idea", "Refutation". If events or news article have many these tags, the tags will support the analysis of concerns. Viewed in supporting debates, citispe@k has similarities to the model of QOC[11] or Deliberatorium[6].

[3]http://www.city.nagoya.jp/shisei/category/53-3-7-4-0-0-0-0-0-0.html

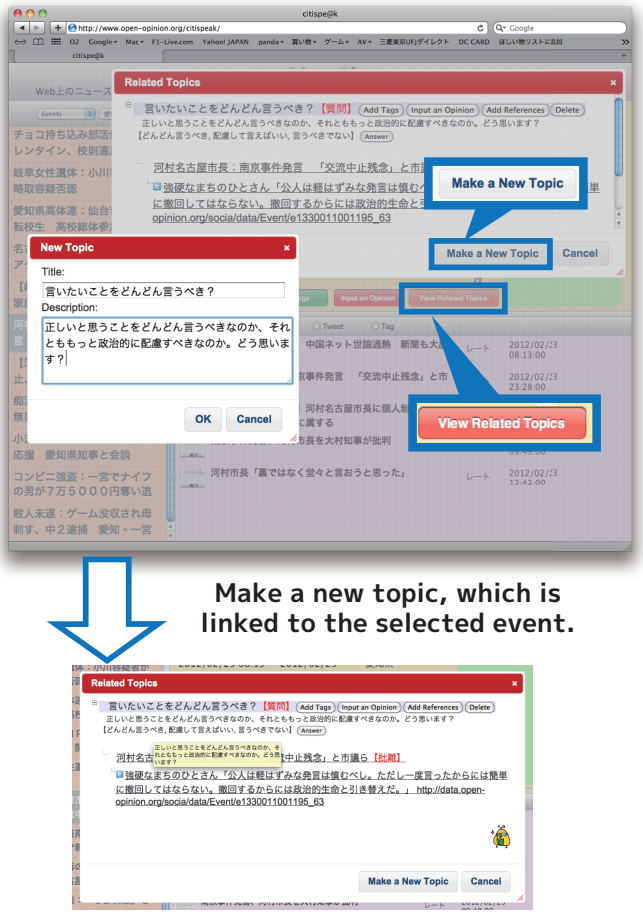

Fig. 7. Make a new discussion topic to the selected event

Fig. 8. Add tags to the selected event

In the QOC model, an issue is structured by theree nodes “Question”, “Option” and “Criteria”. Deliberatorium supports debates or opinion adjustments by classifying opinions into “Agree” or “Disagree” and kinds of opinions. Tagging to opinions are voluntary in citispe@k due to getting many opinions. If tagging were required in citispe@k, users would input little opinions. Users carefree input their opinions with no tag, but the opinions with tags are well conjugated in concern assessment.

Fig. 8 shows the example of tagging to an event. The “Add tags” button will show you the tagging view. In the tagging view, existing tags are listed in the popup menu. Users select appropriate tags from it.

V. Comparison with Other Tools

A. Comparison Criteria

Based on our study, we propose the following criteria driven by effective e-Government and e-Participation requirements for comparing e-Participation support systems for regional communities:

1) Transparency
2) Data Re-Usability
3) Pervasiveness
4) Discussion Appeal
5) Structuring

Transparency is a requirement of all Government 2.0 initiatives. It consists in the open publication of available data by organizations. We extend this definition to include data inherent to the discussion system (comments, discussions). In most cases, transparency is either site-dependent, or making data available in a different form (polls). User-dependent transparency means that a level of privacy exists for which users can conduct private interactions (e.g Twitter). When the transparency is public, this means that all data from the system is made public.

Data Re-Usability relates to the potential of data to be re-used without technology-specific boundaries. As e-Government tools are not limited to the Internet (e.g use of SMS) or one specific application, data that is relevant to the citizens ought to be free from such boundaries so that it can be accessed by various tools and applications. It should also be structured according to the domain of debate, rather than a tool-specific domain. N/A (for non-available information) means that some form of interoperability for data has been suggested but was not found in our tentative use.

Pervasiveness requires the system to be ubiquitous in its accessible interfaces. Here, it describes a system that can be used from any terminal, most notably tablets and mobile phones since they have become the first computing platform and thus a much easier means to reach citizens. This qualifies the capacity of the tool to penetrate everyday use according to technology trends.

Discussion appeal is defined as the capacity of the system to engage its users in proactive discussion by interesting them to a topic at hand. For example, citizens are more eager to

TABLE I
COMPARING O_2 WITH EXISTING TOOLS

e-Participation Tool	Transparency	Data Re-Usability	Pervasiveness	Discussion Appeal	Structuring
Message Boards	Site-dependent	Format-dependent	Mainly Web	Site-dependent	Time-centric
Chatrooms	Site-dependent	Logs	Requires Real-time	Site-dependent	None
Questionnaires	Public Polls	Figures	Medium-dependent	Not proactive	Dependent
SMS Questionnaires	Public Polls	Figures	High	Not proactive	Dependent
Site Comments	Site-dependent	Format-dependent	High	High	Article-centric
Microblogs	User-dependent	API (Twitter)	Very high	High	Hashtags
MIT Deliberatorium	Public (login)	N/A	Mainly Web	Medium	Issue-centric
Cohere	User-dependent	Dataset (planned) [9]	Mainly Web	(out of domain)	Knowledge-centric
O_2	Public	Dataset, Ontology, API	High (mobile-oriented)	High	Automated Event-centric

react to ongoing events and current news than they are to use e-participative involvement tools that require a certain degree of initiative and interest for debate. Thus, from this point of view, capacity of news sites to involve can be for example assessed through the number of comments that readers post on news articles. Particular events often being a catalyst for regional communication and interaction, in this sense regional news sites when they exist are more able to involve than are dedicated e-Participation systems. In this sense, we believe that a system with higher chance of using the swarming effect (see Sec II) has a higher chance of involving citizens, a higher discussion appeal.

Structuring is an important vector of debate, as it helps focus the discussion for every participant. It involves 1. adequate separation of items relevant to different regional stakeholders 2. clustering of similar or related items in an understandable way to avoid redundancy but also provide numerous information sources relatively to one topic of discussion so that bias is avoided. E.g, news aggregation systems provide such clustering, but do not allow commenting on the clusters themselves.

B. Discussion

We conduct a qualitative comparison in Table I between O_2 and a panel of eParticipation discussion support tools: message boards, chatrooms, questionnaires, site comments, microblogs, the MIT Deliberatorium, and Cohere.

The data of message boards, when it is publicly available, is normally highly format-dependent, depending on the message board software used (phpBB, vBulletin, Futaba, Facebook groups, etc), and the view in case one wants to perform page scrapping. Also, its structure is mostly time-centric, as criticized by the MIT Deliberatorium research. The use of swarming to involve users also depends on the message board. A message board addressed to a local community that provides discussion opportunities about current events regularly would have a good capacity to involve citizens, despite its lack of automation and structuring. The lack of automation requires a lot of community management work to find articles, tweets and other background information to enrich threads. The lack of structuring could lead to soapbox and balkanism issues. Finally, message boards are used mainly on the desktop-PC Web and their use is declining overall, in profit of news sites commenting systems and social networks such as microblogs.

Chatrooms are highly site-dependent when it comes to transparency, depending on whether or not logs are made public, and private logs retrieved from users raise the question of trustability. There is also no structuring whatsoever and the data must be sequentially cut. This can be done using previous research we have conducted [12], however although this is useful for discourse analysis, the problem of lack of overall structure remains. Finally, chats may be pervasive depending on the system, but they require real-time involvement, which is an important drawback compared to other systems where discussion can be conducted asynchronously.

Questionnaires and SMS questionnaires, by being highly focused on specific problems, and normally pervasive (especially in the case of SMS questionnaires), are guided discussions that have high capacity for involvement. Flocks can be created although users are not proactive in this form of swarming (if there is no questionnaire, there is no flock). However, they normally have no discourse structure, and their ability to create long-term discourse communities after a flock is subject to doubt. There is appeal to answer and emit opinions, but not specifically to discuss. Finally, data that is made transparent normally consists in public poll figures. In short, questionnaires cannot be considered as a complete eParticipative tool for communities, although they are best as part of one and useful for input probing [13].

Site comments basically suffer from the same drawbacks as message boards in terms of public data and format dependence. Important news sites in Japan require a paying subscription to access archives and their comments. Also, page scrapping can be rendered difficult by the Javascript-controlled display of comments. However, news sites are pervasive since a vast majority propose mobile interfaces and/or applications. They also show a very high swarming effect and discussion appeal on local news that communities feel concerned about, only it is not leveraged to create long-term debate. Finally, the main drawback of site comments is that they are article-centric and centralized on only one news site at a time. There is no possibility to comment a cluster of news as an event, and comments (which constitute basic discourse) have no exploitable structure.

Microblogs are probably the most pervasive, highly used on mobile interfaces by a various panel of citizens, their main representative being Twitter. The swarming effect is also important, as hashtag trends emerge in correlation with important

events in time, and the discussion appeal on microblogs is high. However, as for site comments, there is few leverage of the swarming effect for structured discourse for regional communities, despite on-line formation of Twitter-localized communities. Still, because it is highly pervasive and has a highly exploitable flock, Twitter through its API data re-usability is mined by O_2 in hope to utilize tweets as discussion seeds and contributions.

The MIT Deliberatorium is mainly a desktop Web interface with issue-centric structure. Although in theory it addresses long-term general issues, when we experienced the site in May 2012, present discussion was focused on punctual events. Thus it could exploit swarming, however there is no automation and discussion seeds have to be created manually.

Cohere was added in this comparison because it is probably the closest system to O_2 in that it chooses a strict ontological approach for structure. As Twitter, it has public and hidden data. According to [9], dataset intercompatibility was also planned. Still, it is highly knowledge-centric, focused on ideas and viewpoints. Although it is appealing to share concepts and knowledge, we believe it is slightly off-domain to constitute an eParticipative tool for local communities.

Lastly, we will explain why we believe O_2 is a better fit in the context of involvement of communities and eParticipation according to the criteria. First, O_2/citispe@k focuses on being as transparent as possible, as per instructed for the Government 2.0 initiatives. The data is entirely publicly accessible and reusable through its dataset, ontology and API. Finally, by employing an automated event-centric structure through regional classification and event clustering of automatically mined news articles and tweets, O_2 aims at leveraging the swarming effect. Second, the interface is mobile-oriented (but can be used through a desktop browser as well), which allows it to be more pervasive according to current technology trends. Third and finally, since it is easier to engage citizens through short-term flocks [7] e.g those of service issues or current events, it is fair to assume that engaging debate and discussion through communities on geographically and timely local issues can lead to more involvement on the long term.

VI. Conclusion

We have presented the e-Participation Web platform O_2, for public involvement of citizens from regional communities in dialogue and participative debate through the use of an innovative technological approach. Following the delivery models of e-Government, the scope of this research covers the informing and involving of the citizens in discussion about regional issues. Representatives can then utilize the tool and the data is automatically openly published in a LOD subset we call SOCIA. The innovative point of the platform is to allow building discussion topics for regional communities by directly commenting on news clustered as events and classified automatically by geography. By assessing the goals of an e-Participative system in the context of e-Government and mobility, we proposed relevant comparison criteria for eParticipative tools to be used by regional communities, and then compared O_2 to existing technologies for supporting regional debate among citizens. Upon this qualitative study, we claim that O_2 can constitute a better approach than existing tools of support, through high focus on openness, data re-usability, pervasiveness, discussion appeal and automated event-centric structuring.

Further research directions that are considered include: sentimental analysis on to visualize concern more easily, development of an opinion search engine, deeper structuring of debate through ontology following previous works [5], insight detection on comments based on generative models [14].

Acknowledgment

This work was supported and promoted by the Strategic Information and Communications R&D Promotion Programme (SCOPE), Ministry of Internal Affairs and Communications, Japan.

References

[1] U. N. D. for Public Administration, D. Management, U. N. D. of Economic, and S. Affairs, *United Nations E-government Survey: Leveraging E-government at a Time of Financial and Economic Crisis*. United Nations, 2010.

[2] A. Macintosh, "Characterizing e-participation in policy-making," in *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*. IEEE, 2004, pp. 10–pp.

[3] M. Wimmer, "Ontology for an e-participation virtual resource centre," in *Proceedings of the 1st international conference on Theory and practice of electronic governance*. ACM, 2007, pp. 89–98.

[4] A. Macintosh, S. Coleman, and M. Lalljee, "E-methods for public engagement: helping local authorities communicate with citizens," *Published by Bristol City Council for The Local eDemocracy National Project*, 2005.

[5] R. Hurwitz and J. Mallery, "The Open Meeting: A Web-based system for conferencing and collaboration," in *Proceedings of the Fourth International Conference on The World-Wide Web*. Citeseer, 1995.

[6] "MIT Deliberatorium." [Online]. Available: http://cci.mit.edu/research/deliberatorium.html

[7] D. Della Porta and M. Diani, *Social movements: An introduction*. Wiley-Blackwell, 1999, 2006.

[8] N. Arrivals, M. Friends, and A. People, "e-participation and governance: Widening the net," *Electronic Journal of E-government*, vol. 3, no. 1, pp. 39–48, 2005.

[9] A. D. Liddo and S. B. Shum, "Cohere: A prototype for contested collective intelligence," in *ACM Computer Supported Cooperative Work (CSCW 2010) - Workshop: Collective Intelligence In Organizations - Toward a Research Agenda*, February 2010. [Online]. Available: http://oro.open.ac.uk/19554/

[10] J. D. M. Rennie, J. Teevan, and D. R. Karger, "Tackling the Poor Assumptions of Naive Bayes Text Classifiers," in *In Proceedings of the Twentieth International Conference on Machine Learning*, 2003, pp. 616–623.

[11] A. MacLean, R. M. Young, V. M. E. Bellotti, and T. P. Moran, "Questions, options, and criteria: elements of design space analysis," *Hum.-Comput. Interact.*, vol. 6, no. 3, pp. 201–250, Sep. 1991.

[12] T. Z. T. O. T. S. Shun Shiramatsu, Jun Takasaki and H. G. Okuno, "System for supporting web-based public debate using transcripts of face-to-face meeting," in *Trends in Applied Intelligent Systems, Proceedings of the 23rd. International Conference on Industrial Engineering and Other Applications of Applied Intelligence Systems (IEA/AIE 2010), Part III*, ser. Lecture Notes in Computer Science, vol. 6098. Springer, Jun. 2010, pp. 311–320.

[13] C. W. Phang and A. Kankanhalli, "A framework of ict exploitation for e-participation initiatives," *Commun. ACM*, vol. 51, no. 12, pp. 128–132, Dec. 2008.

[14] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.