WeKnowIt

Emerging, Collective Intelligence for Personal, Organisational and Social Use

FP7-215453

D9.4.2

Final white paper on Collective Intelligence

Dissemination level	Public
Contractual date of delivery	Month 36, 31-3-2011
Actual date of delivery	18-04-2011
Workpackage	WP9 Project Management
Task	T9.4 Quality Management
Type	Report
Approval Status	Approved
Version	2
Number of pages	40
Filename	D9_4_2-man_2011-04-17_v02_certh_Final_White_Paper_on_CI.doc

Abstract

This document is the final version of a public white paper on Collective Intelligence. The state of the art in Collective Intelligence related activities is presented followed by an analysis of the main parts appearing in the Collective Intelligence life-cycle. Several applications implementing Collective Intelligence are then reviewed, based on this analysis and a resulting categorization. Finally, the WeKnowIt contribution towards Collective Intelligence is outlined, along with the WeKnowIt applications that leverage it.

co-funded by the European Union

History

Version	Date	Reason	Revised by
1	12-4-2011	Initial version	S. Diplaris
2	18-4-2011	Final version	Y. Kompatisaris

Author list

Organization	Name	Contact Information
CERTH	Sotiris Diplaris	diplaris@iti.gr
CERTH	Yiannis Kompatsiaris	ikom@iti.gr

Executive Summary

This document is the final version of a public white paper on Collective Intelligence. The state of the art in Collective Intelligence related activities is presented followed by an analysis of the main parts appearing in the Collective Intelligence life-cycle. Several applications implementing Collective Intelligence are then reviewed, based on this analysis and a resulting categorization. Finally, the WeKnowIt contribution towards Collective Intelligence is outlined, along with the WeKnowIt applications that leverage it.

Table of Contents

Table of Figures

1. Introduction

Social media sharing properties, such as Flickr, Facebook and PicasaWeb host billions of images and video, which have been annotated and shared among friends, or published in groups that cover a specific topic of interest. The fact that users annotate and comment on the content in the form of tags, ratings, preferences etc and that these are applied on a daily basis, gives this data source an extremely dynamic nature that reflects events and the evolution of community focus. Although current Web 2.0 applications allow and are based on annotations and feedback by the users, these are not sufficient for extracting this "hidden" knowledge, because they lack clear semantics and it is the combination of visual, textual and social context, which provides the ingredients for a more thorough understanding of social content. Therefore, there is a need for scalable and distributed approaches able to handle the mass amount of available data and generate an optimized 'Intelligence' layer, also called Collective Intelligence, that would enable the exploitation of the knowledge hidden in the user contributed content.

Although Collective Intelligence is at least as old as humans and appears in a wide variety of forms e.g., bacteria, animals, computer networks, it is now occurring in dramatically new forms. For example, Google[1] uses the knowledge millions of people have stored in the World Wide Web to provide useful answers to users' queries and Wikipedia[2] motivates thousands of volunteers around the world to create the world's largest encyclopedia. With new communication technologies and using the Internet as host, a large number of people all over the planet can now work together in ways that were never before possible in the history of humanity.

There already exist a number of approaches, which are based on user-contributed content in order to provide useful information for various applications. For example, mobile location information and uploaded content is used to monitor online traffic and generate traffic patterns in [1], connect citizens in Boston [2], share nature experience [3], discover travel patterns and provide travel advice [4][5], or communicate problems in a city [6]. But what exactly is Collective Intelligence and how can we benefit from it; The MIT Center for Collective Intelligence[3] frames the research question as "How can people and computers be connected so that-collectively-they act more intelligently than any individuals, groups, or computers have ever done before?". It is now more important than ever for us to understand Collective Intelligence at a deep level so as to

[1] http://www.google.com

[2] http://en.wikipedia.org

[3] http://cci.mit.edu

take advantage of these new possibilities. A series of projects aiming at harnessing and using Collective Intelligence is fostered by the MIT Center for Collective Intelligence. These include the Climate Collaboratorium [7], which deals with climatic changes and the Collective Prediction effort, which tries to make accurate predictions about future events such as product sales, political events, and outcomes of medical treatments [8]. Collective Intelligence is also used in healthcare [9], while studies are conducted on its applications in today's organizations [10]. However, the main characteristic of such applications is that they are mostly based on collecting well-structured contributions through specific applications, on shallow statistical processing of the contributions and their visualization. Very few focus on analysis of implicit relations in user-generated content and feedback and on dealing with unstructured large-scale data, where an important source of knowledge is hidden.

This white paper focuses on Collective Intelligence approaches, which are implemented by using media and users relations, actions and interactions in social networking and sharing applications also known as Web 2.0 applications. In these applications, large-scale user contributions and more specifically, tags, which can be used to extract Collective Intelligence, suffer from a number of limitations, such as polysemy, lack of uniformity, and spam, thus not presenting directly an adequate solution to the problem of content organization. Therefore, how to manage, index and search for this content effectively and efficiently is becoming an increasingly important research topic. There have been many approaches dealing with the relevant tasks of tag refinement (to refine the unreliable user-provided tags) and automatic annotation or tagging, especially for user-contributed photos. In [11] the Content-based Annotation Refinement (CBAR) method re-ranks the tags of an image, reserving the top ones as the refined results. A more elaborated method [12] refines and enriches tags based on the visual and semantic consistency residing in social sites. Other algorithms perform automated tagging of an untagged image, either by building classifiers for individual semantic labels [13][14], or by learning relevance models between images and keywords[15][16]. The most frequently used method for automatic tagging is the semi-supervised graph-based level propagation technique [17], where a graph is constructed to model the relationship among individual images in terms of visual similarity. Such approaches, although they make use of the available content collections and their connections, they are trying to improve the quality of each independent user contributed content item and therefore they do not directly address the collective nature of social media and the objective of extracting the hidden Collective Intelligence, which can be used in applications other than annotation and retrieval.

Existing approaches towards extracting Collective Intelligence usually build upon restricted combinations from the available social media attributes. For example, in [18] geo-locations and tag information are used in order to generate representative city maps. In [19] tags and visual information together with geo-location are used for objects (e.g. monuments) and events extraction. Tags from flickr images and timestamp information are used in [20] to form a chronologically ordered set of geographically referenced photos and distinguish locals from tourist travelling. The description of city cores can be derived automatically, by exploiting tag and location information [21]. The approach is able of distinguishing between administrative and vernacular uses of place names, thus avoiding the potential for confusion in the dispatch of emergency services.

But besides these combinations, user generated content can be viewed as a rich multi-modal source of information including attributes such as time, favorites and social connections. For example, beyond harnessing content and the surrounding tags or text, limited effort has been made to include the social patterns into the media analysis. The important aspect of fusion of modalities and different sources is currently lacking in most existing Collective Intelligence applications.

In this whitepaper we discuss the existing approaches towards Collective Intelligence, by focusing on the web-based and social networks approach (Section 2). For this purpose a state-of-the-art survey is given by first defining the different features of the existing Collective Intelligence applications in terms of input sources and analysis dimensions. Then, the Collective Intelligence applications are reviewed and categorized according to the previously defined criteria (Section 3). We also present the WeKnowIt approach towards Collective Intelligence and the methodology upon which the WeKnowIt applications have been built (Section 4). A significant amount of WeKnowIt developed applications are then presented, classified accordingly to the defined categories. The classification of the applications clearly shows the capability of WeKnowIt Collective Intelligence methodology to exploit different input data sources with combined analysis techniques in order to deliver enhanced services to the end users. Finally, the white paper concludes with a discussion on the emerging issues and future challenges in developing Collective Intelligence applications (Section 5).

2. Collective Intelligence approaches

There is not a unique definition of Collective Intelligence (CI). According to [22], CI is the capacity of human communities to evolve towards higher order complexity and harmony, through such innovation mechanisms as differentiation and integration, competition and collaboration. In [23] Collective Intelligence is defined as a shared or group intelligence that emerges from the collaboration and competition of many individuals.

CI appears in a wide variety of forms of consensus decision making in bacteria, animals, humans, and computer networks. The study of CI may properly be considered a subfield of sociology, of business, of computer science, of mass communications and of mass behavior, a field that studies collective behavior from the level of quarks to the level of bacterial, plant, animal, and human societies. Collective Intelligence is the Intelligence which emerges from the collaboration, competition and coordination among individuals. The resulted collective intelligence is an Intelligence greater than the sum of the individuals' intelligence.

In 1996, B. A. Huberman in his article "The social Mind" [24], used the term "Distributed Intelligence" to describe collective behavior giving examples from economy, social insects and the scientific community:

"Intelligence is not restricted to single brains; it also appears in groups, such as insect colonies, social and economic behaviour in human societies, and scientific and professional communities. In all these cases, large numbers of agents capable of local tasks that can be conceived of as computations, engage in collective behaviour which successfully deals with a number of problems that transcend the capacity of any individual to solve. In most instances they do so in the absence of global controls, while exchanging information that is at times inconsistent, often imperfect, and usually delayed."

Computational Collective Intelligence can be understood as an AI sub-field dealing with soft computing methods which enable making group decisions or processing knowledge among autonomous units acting in distributed environments. Web-based systems, social networks and multi-agent systems very often need these tools for working out consistent knowledge states, resolving conflicts and making decisions.

In this sense different computational research fields can benefit from enabling Collective Intelligence, such as:

- **Multi-Agent Systems.** Multi-Agent Systems (MAS) are a computational paradigm, especially developed to conceptualize complex systems. It offers a new way of modelling based on emergence originating from interacting entities. This way the nature of Collective Intelligence can be captured in populations of autonomous individuals.

- **Web-based analysis.** Web 2.0 applications and sites are among the main sources of CI. Web 2.0 concepts have led to the development and evolution of web-based communities, hosted services, and applications; such as social-networking sites, video-sharing sites, wikis, blogs, and folksonomies. In Web 2.0 applications, users play an active role by contributed their own user generated content. Such can be videos and images that is uploaded in multimedia sharing sites such as Flickr, Panoramio, YouTube. This content can be accompanied with user metadata (title, tags, description) or sensor content generated data (e.g. gps location, exif information, data and time that the photo has been taken). The content can also be text (e.g. Wikipedia) that is written in a structured way from many individuals. Forums and blogs are also Collective Intelligence content sources, since user conversations and discussions can be retrieved and analyzed. Such an analysis can provide information about trends, mass evolution, dominant topics for a specific time period, etc. Finally, user feedback is another important source. This information can be retrieved from several sites related with products and services, containing customer/user reviews. Web sites presenting or selling electronics, computers, travel services, books are some examples in which users can comment and rate products and services.
- **Social Networks – based analysis**. The analysis of the structures of social networks reveals information on general properties of the relations, on the hierarchical composition of a network and on the roles and positions of people. Computational Collective Intelligence techniques can be of benefit in creating and supporting social networks, as well as in groupware handling. Also social network dynamics and semantic communication are affected by the use of Collective Intelligence.

The technologies developed within the WeKnowIt project focus on the latter two forms of Collective Intelligence (web-based analysis and SNA). Therefore in this paper we will also restrict our survey to these two research fields.

2.1. Collective Intelligence sources

There are several types of collective intelligence sources. Such can be user generated content (videos, images) that is uploaded in multimedia sharing sites such as Flickr, Panoramio, YouTube. This content can be accompanied with user metadata (title, tags, description) or sensor content generated data (e.g. gps location, exif information, data and time that the photo has been taken). In addition, many implicit and explicit users relations (e.g. friend of) and actions (e.g. "like", favorite, etc) provide useful input. Content can also be text (e.g. Wikipedia) that is written in a structured way from many individuals.

Forums and blogs are also Collective Intelligence content sources, since user conversations and discussions can be retrieved and analyzed. Such an analysis can provide information about trends, mass evolution, dominant topics for a specific time period, etc.

Finally, user feedback is another important source. This information can be retrieved from several sites related with products and services, containing customer/user reviews. Web sites presenting or selling electronics, computers, travel services, books are some examples in which users can comment and rate products and services.

Summarizing, Collective Intelligence input comes from user's conversations and discussions and user-(and sensor) generated content, through media upload, commending, rating, annotating, tagging and co-editing. This input with the use of the Web 2.0 technologies has already provided a type of intelligence in the form of recommendations, customized presentations etc. Nevertheless, using advanced content, tag, statistical and network analysis techniques the intelligence that can be generated can be of much higher level resulting in personalised services, trend detection, social connectivity and ability to extract meaning from content.

Therefore, Collective Intelligence systems can exploit different modalities of input, ranging from single visual, textual, audio or user information, up to the fusion of such media sources, i.e. annotated images, audio or audiovisual content, or even the combination of such content with social networks structure graphs.

2.2. Dimensions of Collective Intelligence applications

Another important aspect characterizing how Collective Intelligence techniques are able to exploit the above data sources is the presence or absence of *spatial* processing. Following current web socializing and multimedia handling user trends, most users nowadays upload, geo-tag and localize their personal photos. Moreover, most Web 2.0 systems make heavy use of localisation of resources in maps. Therefore, many recommendation, presentation or prediction Collective Intelligence techniques are used for enhancing location-based systems and services.

A different dimension, vertical to the *Location* aspect is the absence or presence of processing *temporal* features. This ability directly connects to the event processing capability. When time aspects of the user generated content are considered it is possible to derive information about *situation* or *events*. Based on massive user contributions, Collective Intelligence techniques results could range from the single representation of events, and expand to past/current events detection or even future events prediction.

More elaborate Collective Intelligence applications are able to capture knowledge in the combined *spatio-temporal* dimension. Considering both space and time information, it is possible to depart from mere presentations of geo-location collections and discover knowledge about routes or map areas with particular interest, as well as detect events in the combined spatio-temporal dimension.

To summarize, Collective Intelligence techniques that deal with Web 2.0 - based and social networks content and structure can be categorised with respect to the different dimensions they embrace (spatial, temporal, or spatio-temporal), the different features they consider (visual, audio, textual or their fusion), and their ability to process the underlying social network structure information.

2.3. Collective Intelligence Application Domains

The implicit knowledge derived through Collective Intelligence analysis methods can be consumed in many different application domains. Consumer Social Group and Emergency Response are two relevant examples that have served as case studies in WeKnowIt.

Travel-related web and tourist location-aware services are an obvious market for leveraging the multimedia search and context-centric delivery capabilities. Also, in *Emergency Response* mining user generated data coming from citizens and enabling real-time discovery of events is crucial for effective planning and action in Emergency Response systems.

The *infotainment* sector, and more specifically the cultural event organisation, constitutes a crucial market for European SMEs since it is highly fragmented and holds long-lasting ties with local societies. Local European communities are proud of hosting events, both as a means of attracting tourists, and as vehicles of diffusing culture and arts through the social fabric. The capabilities that Collective Intelligence results provide can have a transformational impact on how events are experienced from the audience, and on the kind of services that are offered to event attendants. Collective Intelligence technologies for analysing massive Web 2.0 content, along with personalisation and content presentation techniques from the Personal Intelligence layer enable the possibility to make use of the project results in events of different scales, from localised ones (e.g. Thessaloniki Film Festival), to events of global reach (e.g. Cannes Festival, Olympic Games, Athens Marathon, etc.).

Brand or *product monitoring* are a further market where the Collective Intelligence capabilities are of great value. Such products can help professional users in the field of marketing as well as individual citizens with given information needs. The availability of new effective means for accessing massively produced web content and of new relevant applications and services will result in wider market opportunities for SMEs with content-centred and Location-Based Service related business models. The technological innovations stemming form Collective Intelligence can possibly help to provide new business opportunities for SMEs in the areas of event organisation, event-related and location-based marketing and providers of information access technologies.

Another application domain for Collective Intelligence is *e-government*. Strengthening citizens' participation in governmental and local societies decision making processes is not only a cornerstone for the notion of democracy, but also the only way for people to influence the decisions and policies that affect their daily lives. Web 2.0 based tools can allow citizens to voice their opinions, stimulate dialogue and for their views/concerns to be taken into account in governmental policies. A major drawback of current relevant applications is that they do not fully support intelligent processing and management of such user contributed information. Thus, both citizens and local decision makers fail to access the large amount of content efficiently and cannot exploit the underlying, hidden knowledge. Collective Intelligence techniques can be applied on this content to extract patterns and knowledge arising, and contribute to a policy planning which reflects better the citizens' needs. Large-scale processing which takes in account content (e.g. images, text, videos), tags, social information (e.g. friendships and citizens communities), location and spatial information can be exploited in order to produce Collective Intelligence, thus providing added-value to the available content and making existing procedures and workflows more efficient. For example, by jointly analysing blog comments and the location of the users submitting the comments, a specific topic (i.e. "bicycle lanes") might arise as dominant for a specific area showing the interest of the citizens on this topic and allowing efficient planning. Moreover, analysis of user-contributed content via Collective Intelligence techniques allows linking with existing structured or semi-structured sources, e.g. by creating Linked Data.

Furthermore, again in terms of Collective Intelligence technologies, Linked Data can be further combined (mashed-up) with any other piece of Linked Data, promoting the usage of the acquired data in various areas. For example, government data on health care expenditures for a given geographical area can be combined with other data about the characteristics of the population of that region in order to assess effectiveness of the government programs. A combination of web 2.0 and semantic technologies can make it possible both for citizens to track legal procedures, understand technical documents, express views, and for elected representatives to better handle the gathered information and transfer it to other bodies as well. Specific e-Government fields that can benefit from such a Collective Intelligence framework are opinion mining aiming at topic and trends detection, public consultation procedures, or problem reporting at local (neighbourhood) level.

Several scientific domains are also impacted by Collective Intelligence technologies. *Computational sociology* is a branch of sociology that uses computationally-intensive methods to analyze and model social phenomena [49]. Using computer simulations, artificial intelligence, complex statistical methods, and new analytic approaches like social network analysis, computational sociology develops and tests theories of complex social processes through bottom-up modeling of social interactions. Collective Intelligence techniques that enable the combination of SNA technologies with media and mass content analysis methods can boost research also in computational sociology.

Finally, the scientific and technological progress resulting from the exploitation of Collective Intelligence will have significant influence on research communities in the topics of Data Mining and Knowledge Discovery, Multimedia Information Retrieval, Contextual Media analysis and fusion, Personalisation and Recommendation and Human Computer Interaction.

3. Collective Intelligence Applications

In this section we will review some of the Collective Intelligence applications that take into account large-scale semi-structured user generated content and apply computational methods in discovering implicit relations and therefore the “hidden” knowledge. We also attempt to classify them according to the categorization and the features these techniques consider, as discussed in the previous section.

Although still at the level of research, there are several applications that exploit the knowledge extracted from massive user contributions. For instance, it is common to derive community-based views of networks, i.e. networks of which the nodes correspond to the identified communities of the original networks and the edges to the relations between the communities. Such views are more succinct and informative than the original networks. It is for this reason that community detection has found applications in the field of recommendation systems [25][26][27][28], as well as for representing user profiles [29][30]. Other applications that make use of the knowledge extracted from tag communities include sense disambiguation [31] and ontology evolution/population [26].

The patterns emerging often show deep interconnections with various world events [32][33] in a way that the evolving world model is captured at each instant.

A leap in the exploitation of user contributed content research is the work in [34] which explores global trends and sentiments that can be drawn by analyzing the sharing patterns of uploaded and downloaded social multimedia. Taking into account both spatial and temporal aspects of content item views and uploads in social media sites and aggregating them, the authors are trying to forecast future events impacting politics, economics and marketing.

In [35] the authors are using human sensors to detect real world events, and generate situation awareness. They describe how spatio-temporal-thematic data in various social media can be aggregated into ‘social pixels’. They derive an image like representation which allows for sophisticated data processing, offering to the user a query algebra tool for posing sophisticated situational queries. The technique is applied in business analysis, seasonal characteristics analysis and political event analytics.

Further, other work shows that the actions of individual Internet users, when properly pooled, can indicate macro trends. There are studies using Search Engine queries for influenza Internet surveillance [36], such as Google Trends [37], search advertisement click through [38], Yahoo search queries [39] and health website access logs [40]. Specifically in [37], Google search engine queries and data from the Centers for Disease Control (CDC) are used to find 45 specific search terms that are related to the percentage of influenza related physician visits. This model allows for monitoring influenza rates 1-2 weeks ahead of the CDC reports.

Table 1 presents a categorization of the above techniques and applications, as well as the CI techniques presented in the Introduction, with respect to the parameters discussed in Section 2.

Table 1: Categorisation of Collective Intelligence applications.

	Data sources				Analysis Dimensions		
Approach	**Visual**	**Textual**	**Audio**	**Users**	**Spatial**	**Spatio-temporal**	**Events/situation**
[25]		+		+			
[26]		+		+			+
[27]		+		+			
[28]		+		+			
[29]		+		+			
[30]		+		+			
[31]		+		+			
[32]		+		+		+	+
[33]		+		+		+	+
[34]	+	+		+		+	+
[35]	+	+	+			+	+
[36]		+					+
[37]		+					+
[38]		+					+
[39]		+					+
[40]		+		+			+
[18]	+	+			+		+
[19]	+	+			+		+
[20]		+		+		+	+
[21]		+			+		+

It can be observed that most Collective Intelligence applications do not employ the spatial or spatio-temporal dimension, instead they mostly make use of textual user generated content in order to provide recommendations, or represent events and situations. However, there have also been developed some more advanced techniques and applications [32][33][34][35][18][19][20] that can process more kinds of input modalities enabling the spatio-temporal and situational dimension.

4. WeKnowIt Approach

4.1. Intelligence Layers in WeKnowIt

The WeKnowIt approach towards Collective Intelligence builds on two aspects: mass content availability provided by a lot of users and availability of analysis techniques and results from different layers. Collective Intelligence methods can be classified based both on the number of the input modalities or layers that they employ in the analysis and the number of users contributing to the data. More specifically, the different Intelligence layers that contribute to Collective Intelligence can be classified to the following:

Media intelligence is the intelligence originated from digital content items (images, video, audio, text) and contextual information analysis, either provided by the user or pre-existing, and their merging. For this purpose, intelligent, automated content analysis techniques are used for different media to extract knowledge from the content itself. Since the amount of data is large and noisy, machine learning, data mining and information retrieval methods are be used. Also the methods are able to fuse information from different sources/modalities, contextual information (e.g. time, location, and EXIF metadata), personal context (profile, preferences, etc.) and social context (tagging, ratings, group profiles, relevant content collections etc.).

Mass intelligence analyzes user feedback. Mass analysis enables input information clustering and ranking as well as information and event categorization. Also, bursts of information can be detected that may indicate potential events (emergency) and trend analysis and prediction. Facts and trends are recognized and modelled by interpreting user feedback on a large scale. For instance, a single road being blocked in a storm may not be very critical, but all access roads being blocked towards a hospital centre may be very critical in the case of an emergency.

Social Intelligence is the exploitation of information about the social relations between members of a community. Nearly everything humans do, they do in a social context because they communicate, collaborate or in some other way interact with other people. Information about the various types of social relations may be represented in communication networks, friendship networks or organization charts.

Organisational Intelligence allows support of decision making through workflows exploiting the generated knowledge and taking into account existing procedures within an organisation. This is quite a departure from traditional methods where knowledge is produced by the individual knowledge worker and collected and integrated manually in knowledge based systems or organisational repositories.

The above classification scheme is not used to categorize completely different approaches; instead it is used as a means to describe how the synergy effect of Collective Intelligence can be of benefit to the end users.

4.2. WeKnowIt Collective Intelligence methodology

The herein described Collective Intelligence techniques in most cases exploit links, references and relations among different content items contributed by the users, thus differentiating from the legacy large scale data analysis techniques. Typical examples of such techniques are Flickr-based visual analysis, tag clusters extraction from massive user tagging and social media-based community detection methods. The integration of such different techniques originating from different intelligence layers could potentially leverage Collective Intelligence within diverse usage scenarios.

However, the Collective Intelligence approach proposed in WeKnowIt moves a step further; instead of a mere concatenation of the different layers intelligent methods, it imposes a pairwise combination of different intelligence layers within the architecture of some of the developed techniques. Multi-modal analysis is often exploited to enhance the results in each intelligence layer. For example, Mass Intelligence tag clustering results are improved by using Media Intelligence visual analysis features when building graph clusters. As a result the produced clusters are evaluated as more coherent, since they incorporate cross-domain knowledge.

Furthermore, the added value of Collective Intelligence is also evident in the integration level, where the different techniques are combined to produce better results in each case. For instance, geo-tagging through visual and tag analysis yields better localisation results, while Flickr annotations can be improved through automatic localisation and tag recommendations emerging from massive user generated content or the user's social network.

Figure 1 depicts the proposed Collective Intelligence approach. User generated content items can be analysed by Media Intelligence techniques, resulting in knowledge enhancements about them, as is the case for deriving routes from spatio-temporal analysis using the VIRaL tool. In a higher level, Mass Intelligence techniques are able of analysing massive Web 2.0 users feedback, aiming at extracting implicit knowledge about more generic situations, as is the case with ClustTour and its ability to identify interesting time periods or areas in cities. In overall, the WeKnowIt Collective Intelligence methodology is able to produce enhanced results by:

- exploiting large-scale user contributed content
- combining different layers in building Collective Intelligence techniques
- fusing results from different intelligent layers

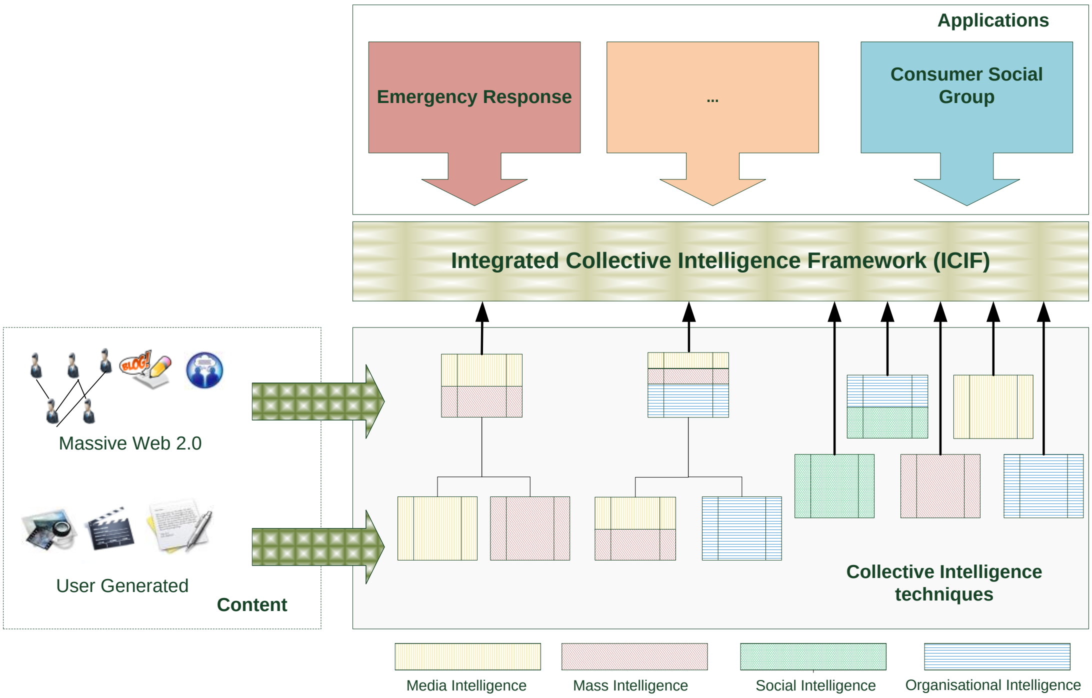

Figure 1: Collective Intelligence methodology in WeKnowIt.

4.3. WeKnowIt Collective Intelligence Applications

In the following we present some representative WeKnowIt applications, which are typical examples demonstrating the WeKnowIt Collective Intelligence methodology.

4.3.1. ClustTour

ClustTour is a WeKnowIt Collective Intelligence application which uses photo clusters corresponding to landmarks and events in order to assist the online exploration of a city. ClustTour provides a means of alleviating information overload, by providing a cluster-based view over very large collections of geotagged photos. By using graph-based techniques to analyse massive user generated content, photos are organized into clusters, and photo clusters are classified as landmarks and events. Subsequently, ClustTour uses a map interface to visualize clusters and to enable users to navigate through them. The application offers two modes of exploration: a city view depicting a high-level view of the most important clusters in the city (ranked by number of photos contained in them) and a Point of Interest (PoI) view centered on the selected PoI and showing landmarks and events in its vicinity.

ClustTour has been extended [41] in order to provide a spatio-temporal content organization layer on top of the photo clusters, and at the same time refine the detected clusters by taking into account the spatio-temporal distribution of photos in the dataset. In that way, ClustTour endows users with enhanced content exploration and browsing capabilities, and at the same time, it improves upon the quality of the presented clusters. In the end, users will get informative views over the interesting spots and areas in a city depending on the temporal context that is of interest to them.

With its ability to analyse massive visual and textual information from user generated content (Flickr) and to present both landmarks and events in a spatio-temporal dimension, ClustTour is a typical case of Collective Intelligence application enabling the analysis of multi-modal input in the large scale.

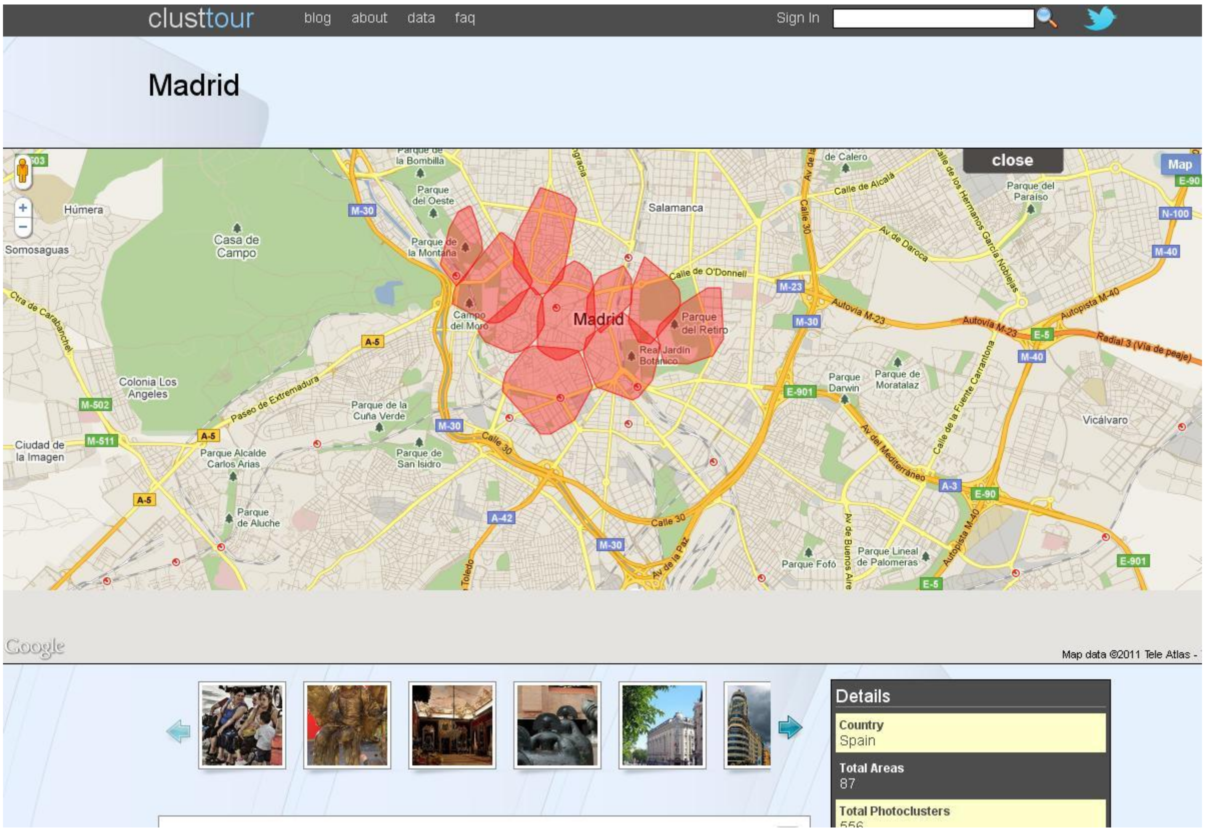

Figure 2: Areas clustering in Madrid with ClustTour.

4.3.2. VIRaL

VIRaL [42] tool is a web-based tool developed within WeKnowIt which is used to identify and localise similar multimedia content under different viewpoints, applicable to any functionality that involves a still image search and retrieval task. The main research principle of VIRaL exploits the fact that typical metadata usually contain a free text description together with some representative user-generated tags or geo-coordinates.

Initially, in order to represent the visual content of any given digital still image, a set of interest points is selected and visual features are extracted locally from their surrounding area. Then a visual vocabulary is defined by clustering methods in such a way that parts of images could be meaningfully assigned to visual words. Then for a new query, its points have to be assigned to the closest visual words of the vocabulary. After this process, two images are considered to be similar if their points are assigned to similar visual words.

When a user query reaches the system the most similar images are returned. Finally, regarding geo-tag estimation, if a user issues a query containing a landmark image against a large database of geo-tagged images, then most probably the top-retrieved results will contain the actual landmark that the query image depicts.

A recent extension to the tool functionality enables users to track routes of friends uploading images in Flickr using spatio-temporal features available in the images metadata.

VIRaL is a Collective Intelligence application that exploits in the spatio-temporal dimension both visual and textual content uploaded by millions of users and available through Flickr.

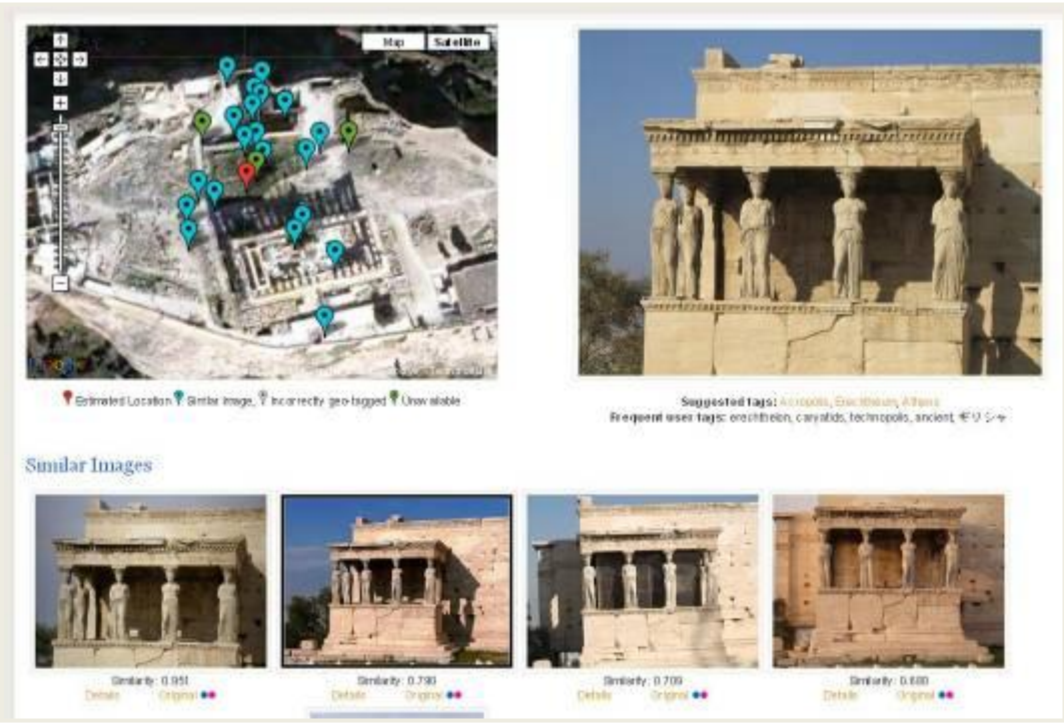

Figure 3: Photo matching and localisation with VIRaL tool.

4.3.3. Travel Preparation tool

The Travel Preparation tool [43] is part of the Consumer Social Group case study of WeKnowIt. It is a web application which provides the relevant information that users need to prepare their travel, e.g. information about the locations, multimedia content, and points of interest. This information is aggregated with content coming from different sources (e.g. Wikipedia), is ranked according to trends to add value to user's experience, and is presented to the user by the use of Mass Intelligence clustering techniques. Morevoer the users are enabled to store results for future visits in online environments visible by his social network friends (delicious).

Figure 4 depicts a snapshot from the online available travel preparation tool.

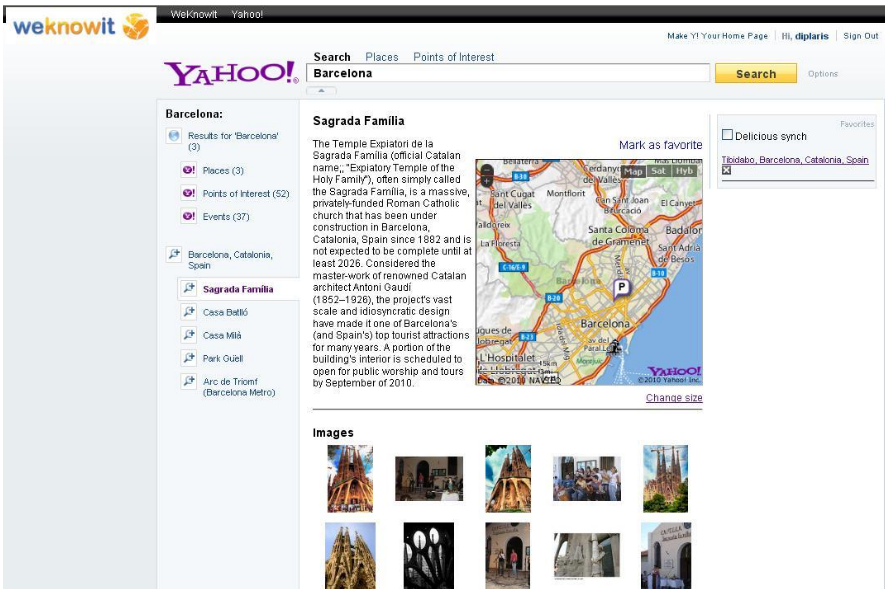

Figure 4: Travel preparation tool.

The travel preparation tool exploits mainly textual information in order to aggregate heterogeneous information and create a complete location context to the end user and his social network. Furthermore, events are supported through event detection techniques that are applied to Wikipedia content.

4.3.4. Mobile Guidance

The WeKnowIt mobile guidance application [43] is also part of its Consumer Social Group use case. It provides guidance to a user during one-day cultural trips. With the features offered by mobile devices and the developed application, users are able to access relevant information about their physical environment and search for new events or points of interest. They are also able of using Social Intelligence techniques to find the position of their social connections, or even notify them in case of emergency. Users can also take pictures of the places they are visiting, which are likely to correspond to the ones chosen in the travel preparation phase. Moreover they can automatically get personalised recommendations about POIs or events based on their profiles and their previous behaviour, as well as recommendations based on massive user content contribution in Flickr and Wikipedia.

Mobile guidance is a mobile prototype that integrates many services and techniques developed in WeKnowIt coming from different Intelligence Layers. Around ten different analysis techniques are employed by the prototype, processing textual, visual and social network structures as input sources, and supporting spatio-temporal features and events.

Figure 5: Mobile guidance - POIs around you.

Figure 6: Mobile guidance - Events around you.

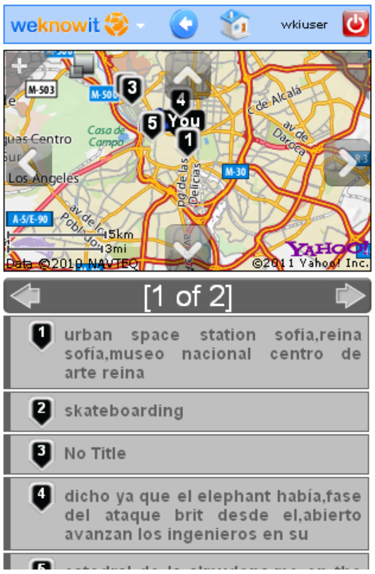

Figure 7: Mobile guidance - Clusters around you.

4.3.5. Fannr

Fannr (Flickr Annotator) [43] is a WeKnowIt application for assisting users in annotating their Flickr photos. It integrates several services implemented in WeKnowIt. It provides two types of annotation support when annotating a single photo, i.e. a) tag recommendation, where the user is presented with a list of tags that they could consider adding to their photo, and b) location recommendation, where the user is presented with several potential locations as to help them remember where the photo was taken.

Both the tag-recommendation and the location recommendation is based on data from different social levels: the user's own photos, the photos of the user's contacts (acquired through social intelligence techniques), and everybody's photos (discovered with mass intelligence methods). Another source of content is the derivation of visually similar photos through the visual analysis geo-localisation methods that are also used in the VIRaL tool (Media Intelligence). Furthermore, for photos that have been localized the tag-recommendation is also based on geographically close-by photos.

When the user has identified tags that they want to add to their photos or identified the location where the photo was taken, the Fannr application also updates the annotation on the Flickr website using the Flickr API.

This is a very indicative example of how Collective Intelligence can enrich user and content items knowledge by integrating different content sources (visual, textual) in the spatial dimension, and by combining analysis techniques from different Intelligence Layers.

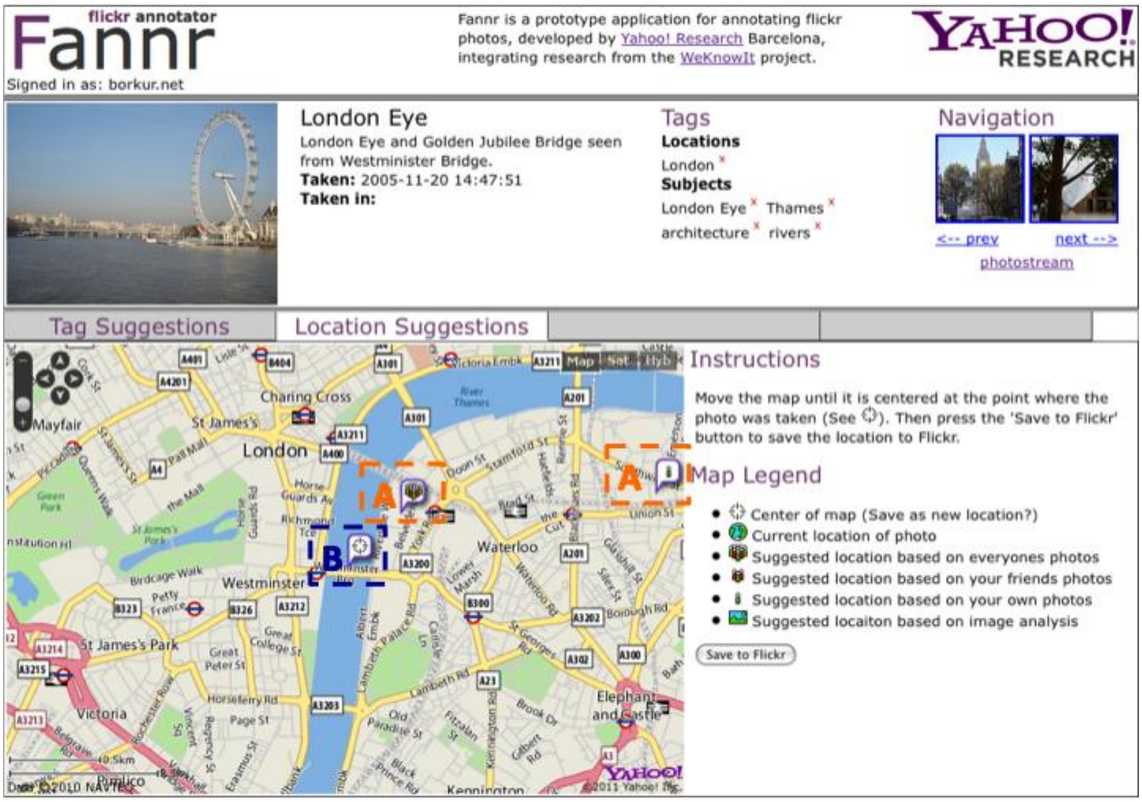

Figure 8: Location suggestion part of Fannr tool.

4.3.6. Reputation System

The WeKnowIt Reputation System describes a way to formalize the assessment of user generated content through explicit reputation. We consider reputation as a quantitative value that explicitly represents a property of an entity of interest, like users and content. The purpose of this reputation system is to assign a reputation to these entities, which has a correlation with a non-visible property.

In particular a score that represents the likeliness of a text of being spam (a service provided by Mass Intelligence layer) and therefore unreliable for ER purposes can be used to determine the reliability of this content especially in the absence of other information. If ratings for content items have been provided by other users (collected via the Personal Intelligence layer) this information will be the basis for calculating the content item reliability. Indication for user reliability is provided by the users social connectivity (determined by Social Intelligence services) under the assumption that the more connected a user is, the more reliable she might be. A strong indicator for a reliable user is a membership in a group of users that is known in advance to be reliable, like all members of the fire department. This functionality comes from the Organisational Intelligence layer. Both content item reliability and user reliability can influence each other. Under the assumption that a reliable user creates reliable content in absence of other information the reliability of the content creator is an indicator of the contents reliability. On the other side when a user creates content that is proven to be reliable this is a strong indication for the user to be reliable herself.

The WeKnowIt Reputation System combines functionalities provided by different modalities and Intelligence Layers, thus it is a solid example of Collective Intelligence. It processes textual and user information provided massively by users and can be applied to characterize many types of content items, such as images, text, audio, as well as the users themselves.

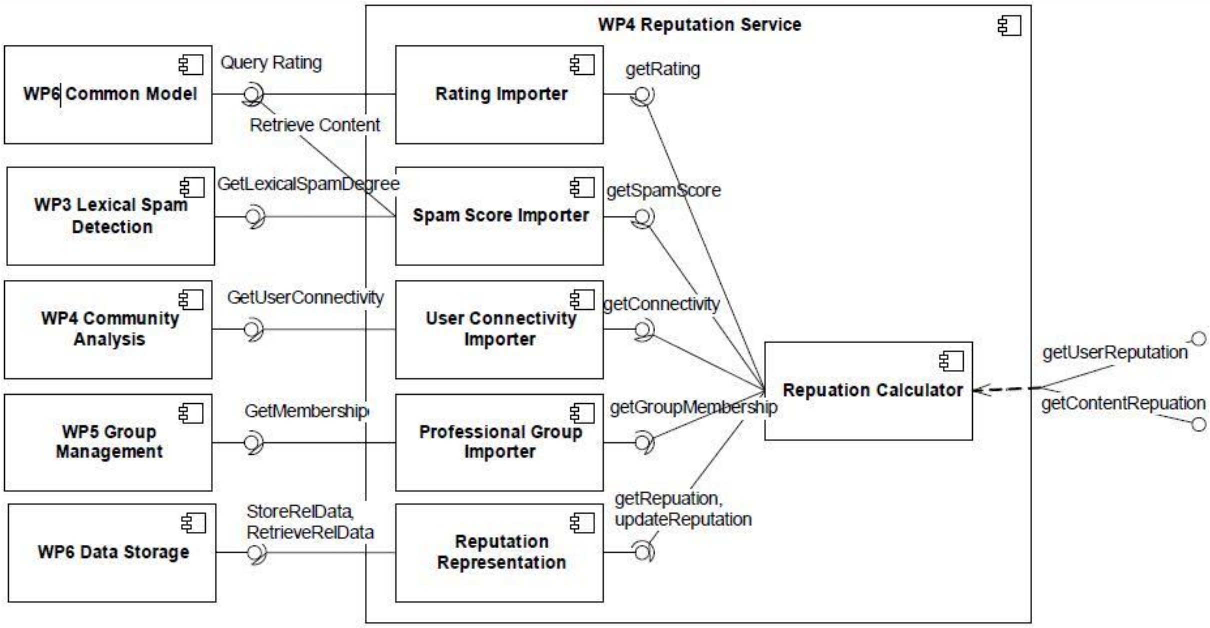

Figure 9: Reputation system architecture.

4.3.7. STEVIE

WeKnowIt's novel mobile application STEVIE [44] enables its users to collaboratively create, share, and modify semantic points of interest (POI). Semantic POIs describe geographic places with explicit semantic properties of a collaboratively created ontology. As the ontology includes multiple subclassifications and instantiations and as it links to DBpedia, the richness of annotation goes far beyond mere textual annotations such as text. With the intuitive interface of STEVIE, users can easily create, delete, and modify their POIs and those shared by others. Thereby, the users adapt the structure of the ontology underlying the semantic annotations of the POIs.

Data mining techniques are employed to cluster and thus improve the quality of the collaboratively created POIs. The semantic POIs and collaborative POI ontology are published as Linked Open Data. STEVIE proposes a new approach for mobile visualization and interaction of temporal information by integrating support for time with today's most prevalent visualization of spatial information, the map. This approach allows for an easy and precise selection of the time that is of interest and provides immediate feedback to the users when interacting with it. It has been developed in an evolutionary process gaining formative feedback from end users.

4.3.8. Emergency Alert

The WeKnowIt Emergency Alert Service [45] has been designed and developed to break the bystander effect and alert contacts of a victim's social sphere in parallel to public authorities. According to the bystander effect, the more bystanders are present in an emergency event, the less likely the victim will be helped; the presence of others inhibits a bystander helping.

The Emergency Alert service is a location-based service that activates nearby members of the social group of a victim in the case of an emergency. It is a mobile service which can be installed as an applet on smart phones (e.g. Android based phones). When activated it works as an emergency call agent and informs social contacts (friends, family, and colleagues) and public authorities about the emergency situation. Current geo-position and routing information is provided. The service is designed privacy-aware and dynamic. Due to its design it can be used in different scenarios, such as the WeKnowIt emergency response and travel scenarios.

For discovering the most likely person to help in the victim's social network, his social network has to be known and possible helpers identified. The network is either built from social interaction data from e-mail, sms, phone, and mobiles where the number of interactions is taken as an indication of social nearness or from social web sites as for example Facebook or Xing.

The Emergency Alert service is another example of harnessing Collective Intelligence from social networks in order to inform about an event or situation. It is able of processing textual and audio content items as well as social interactions. Also, it enables the spatial dimension by providing routes and geo-locations.

4.3.9. Multi-modal recommendation engine

The WeKnowIt Recommendation Service (RecS) has been designed to reveal connections between users, based on their activities in the WeKnowIt platform. Its backbone is a multi-modal clustering process which identifies clusters of resources in different feature spaces [46]. The main difference to the typical recommendation algorithms is that it differentiates between users' interconnections and assigns weights on each kind of similarity they share i.e. social similarity (through their social networks), content similarity (through tags and photos). To estimate the weights, the service makes usage of intelligence provided by Media, Mass and Social Intelligence and combines them so that each cluster contains resources that are connected based on one or more kinds of intelligence. For example, one cluster may contain visual-alike images, whereas another cluster may contain images with related tags.

The recommendation service employs multiple kinds of intelligence together, each of which pertains to a different modality of the resources by exploiting the cross-Usage of Intelligence. In this sense, both visual and textual information is processed. It also takes into account the underlying social networks structure, since its final goal is to improve the community knowledge by exporting groups of users that are interested in the same topic as derived by their activity in the WeKnowIt platform and their uploaded content.

4.3.10. Image recognizer

The WeKnowIt Image Recognizer mobile application [43] provides the user with detailed information on the location and name of a Point of Interest (POI) that she took picture of. The application uses services developed in WeKnowIt in order to recognize a Point of Interest (POI) on a picture and determine its geo-location. Additionally, the application consults Wikipedia to obtain the description of a discovered POI and presents a map centred on the discovered POI using the built-in mobile map software or Google Maps.

The Image Recognizer mobile application combines the results of Media and Mass Intelligence layers. Specifically, the geo-location of a POI is discovered using the same visual analysis algorithms implemented in the WeKnowIt VIRaL tool, whereas the recognition of the POI is the output of a WeKnowIt Mass Intelligence method which returns a ranked list of Points of Interest for a particular geo coordinate. The service processes both visual and textual user generated content sources and mainly exploits the spatial dimension.

Figure 10: Image Recognizer application response.

4.3.11. Emergency Response prototype

The Emergency Response prototype [47] is an application comprised both by mobile and desktop versions, that facilitates Emergency Response personnel and citizens to manage emergency situations.

The Emergency Response prototype has four primary aims. It enables individuals to upload visual textual and audio information to the system about an emergency incident using mobile or desktop devices; this information is enriched through the action of several Intelligence Layers present in WeKnowIt. This information is then presented to citizens and ER personnel allowing them to understand the incident and make improved decisions on the basis of this information. Additionally it allows the ER to perform post-incident analysis using a selection of tools.

In an Emergency Response (ER) scenario, upon an emergency event (e.g. fire, flood, etc.) a user logins to the application and is able of capturing the event and contextualize it with metadata, e.g. tags. Users may be citizens or ER personnel. Such user groups are handled through Social Intelligence SNA techniques. Mass Intelligence techniques are then employed in order to detect spammy content or automatically suggest further ER related tags (from an ER domain ontology) to the user for annotating the captured content item.

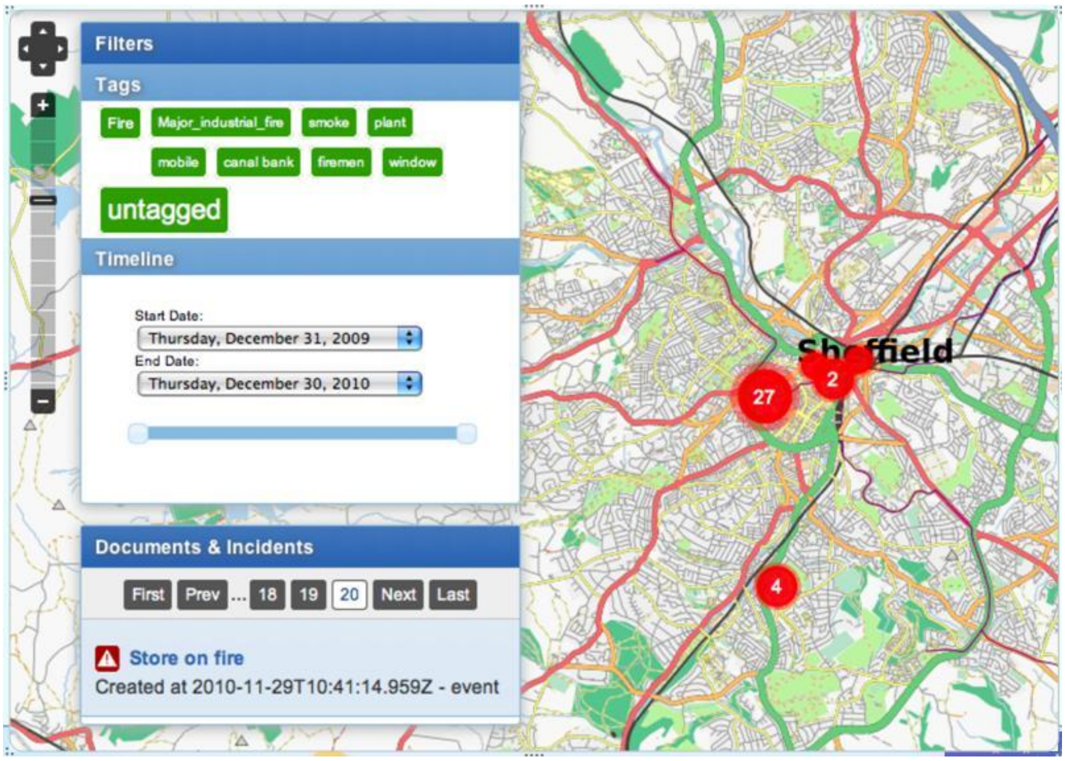

Figure 11: Clustering interface of WeKnowIt ER prototype.

In the Emergency Responders side, Mass Intelligence techniques are applied in the user contributions, which are also enriched by relevant content from Web 2.0 sources. A community detection technique is applied on the received tag volume, so as to induce clusters of the received image data. The ER personnel can monitor the situation via visualization of the incoming information on a map. They are able to automatically or manually filter the incoming information about content items using tag filtering or another Personal Intelligence technique that filters the content items on the map based on the recent pointing and clicking activity of the user on the screen. Meanwhile, if a user of the application is in danger, she is able of activating her connections in her social group, by utilizing the Social Intelligence Emergency Alert service. Finally events and user interactions are represented by means of the WeKnowIt core ontology (namely CURIO [48]).

In overall, the different analysis layers that are used in the different stages of the event contribute altogether to the leveraging of Collective Intelligence, which in turn yields better handling of such Emergency Response events.

4.3.12. Applications Categorization

A categorisation of the WeKnowIt applications presented in the above subsections is given in Table 2, according to the data sources and analysis dimensions discussed in Section 2.

Table 2: Categorisation of the WeKnowIt Collective Intelligence applications.

	Data sources				Analysis Dimensions		
Approaches	**Visual**	**Textual**	**Audio**	**Users**	**Spatial**	**Spatio-temporal**	**Events/situation**
ClustTour	+	+		+		+	+
VIRaL	+	+		+		+	
Travel preparation tool	+	+		+	+		+
Mobile Guidance	+	+		+	+		+
Fannr	+	+		+	+		
Reputation System		+		+			+
STEVIE		+		+	+		+
Emergency Alert		+		+	+		+
Multimodal recommendation	+	+		+			
Image Recognizer	+	+			+		
Emergency Response prototype	+	+	+	+		+	+

It is evident that the WeKnowIt Collective Intelligence applications exploit in a higher degree the spatial, situational and spatio-temporal aspects in analysing massive user generated content, by using algorithms that combine multi-modal input sources, in comparison with the majority of the state-of-the-art Collective Intelligence applications. The WeKnowIt methodology, as integrated in the WeKnowIt platform, is built in a manner that strengthens cross-usage of intelligence and can support such advanced social media applications.

5. Conclusions - Challenges

Central to web 2.0 is the concept of user-contributed content (i.e., users providing keywords to describe digital data sources, express their preference, etc). The process of having end-users adding their own metadata to internet resources, introduces a new way of digital data sources' organization in the web that constitutes the core process in a number of web 2.0 applications. This activity is completely subjective and does not rely on a controlled vocabulary of a pre-defined structure. This fact, although positive from the perspective of dynamically capturing the community's point of view and general trends at a given time, causes the contributed content to be un-structured and lacking of clear semantics.

The herein reviewed Collective Intelligence techniques are trying to harness one or more forms of such online user contributions in order to benefit end-users by employing large-scale recommendation, prediction, detection, representation or summarization analysis techniques. However, such advanced efforts are still very limited, compared to the widespread of social media usage.

WeKnowIt results impact citizens, companies and public organizations. Its applications automatically provide to the users access to information coming from different sources that would be otherwise unreachable within reasonable time in real-life situations. Also, companies are enabled to exploit trends in communities and access to new markets and create new information based services. Increased reaction rate in emergency situations has as a result saving human lives.

The WeKnowIt Collective Intelligence methodology is able to transform the large-scale and poorly structured Social Media to a Social Sensor of meaningful topics, entities, points of interest, social connections and events. The approach towards Collective Intelligence combines the exploitation of massively available content in web 2.0 sources with a multitude of research methods coming from different intelligence layers. These are combined at different levels and with various techniques in order to produce additional, higher-level information, more reliable, by properly exploiting the different input sources.

As a consequence a number of issues and challenges arise in this framework:

- **Wrong, noisy data**. User contributed content is very noisy containing many non-relevant contributions either intentionally (spamming) or unintentionally. The lack of constraints with respect to tagging is the source of numerous annotation quality problems, such as spam, misspellings, and ambiguity of semantics. Moreover, the lack of (hierarchical) structure in the contributed information results in tag ambiguity (a tag may have many senses), tag synonymy (two different tags may have the same meaning) and granularity variation (users do not use the same description level, when they refer to a concept). A related challenge stems from the fact that there is currently a variety of metadata associated with online content items; for instance, a photo can be described by a title, a set of tags, and GPS coordinates. However, not all photos consistently contain all of these metadata. Therefore, it is hard to devise sufficiently resilient knowledge discovery methods given that metadata is incomplete or of dubious quality. In WeKnowIt, approaches like content pre-processing and cleaning and spam detection are applied before analysis.

- **Efficiency of semantics and analysis**. In many cases the performance of existing algorithms is not enough for Collective Intelligence related applications, since most existing algorithms are suitable for running in medium-to-large scale. In these cases, new algorithms should be improved or existing ones should be adapted, in order to be applicable in large scale.
- **Performance, scalability (speed, storage, power)**. Scalability is an important issue since as it has been described, Collective Intelligence is based on mass amount of data. This huge volume of user generated data raises scalability issues that significantly compromise the performance (in terms of accuracy) of algorithms operating on such data. The situation gets worse in cases where the employed algorithms aim at extracting knowledge patterns that only becomes stable after a specific usage period. In these cases the algorithmic output can be considered meaningful only when the total amount of processed data is huge, which poses very demanding requirements in terms of algorithmic design, computational power, memory and storage. This includes parallelization and distributed techniques, speed, storage, memory and relevant considerations, which are indispensable in order to advance to real-time systems offering Collective Intelligence capabilities.
- **Fusion and Integration**. One of the most outstanding features for social media is their intrinsic multimodal nature that opens-up new opportunities for content analysis and consumption. Different types of information are associated with the same digital resource expressing different aspects of its meaning (e.g., textual and visual descriptions, spatio-temporal information, popularity, etc). For instance, an image in Flickr is associated with the tags that have been assigned to it, the users that seem to like it and mark it as favorite, the visual features that describe the visual content of the image, and possibly spatial or temporal information that denote the spatial and temporal context of this particular image. Even though all these facets of information are not combined naturally with each other, still they carry knowledge about the resource, with each facet providing a representation of the particular resource in a different

feature space. However, the heterogeneous nature of multi-source information raises serious obstacles for its efficient exploitation. Common representation and normalization approaches are needed, while it is not clear whether a common framework and approach can be applicable in all cases. Multi-modal fusion techniques are particularly motivated by this fact, aiming to discover otherwise inaccessible information.

- **User Interfaces** with specific functionalities must be developed in order to support user interaction both for contributing the content for the generation of Collective Intelligence and for consuming the results. This is the objective of Personal Intelligence layer in WeKnowIt.
- **Trust, security, privacy** is always a concern for users when they contribute content and especially when this is further analyzed by the application. It should be made clear to the users that their contribution remains anonymous and/or is used for an objective they consent (e.g. emergency response).
- **Applications and commercialization**. Integration with services - organizations. It is not straightforward how the results for Collective Intelligence generation can be integrated into existing business and organisation procedures.

In overall, we believe that WeKnowIt has opened a new way for harnessing the power of the users' public contributions. We expect it to serve as a paradigm of a new generation of social media applications that will use massive users intelligence as an integral component, by taking advantage of everyday human online activities. Future work in extending our Collective Intelligence framework is tailored around the following axes:

- **Real Time** Collective Intelligence. The development of approaches towards providing knowledge derived via harnessing Collective Intelligence in real-time environments is a strategic aim. The means to achieve this is the employment of distributed and parallel methods in analysing the massive content in the large scale.
- **Aggregation** of multiple input sources. The analysis of multiple Web 2.0 sources by single Collective Intelligence techniques is a step beyond our existing approaches. For example input from Twitter and Flicker can be analysed together by the same Collective Intelligence algorithm. Mash-ups are a type of application that can fuse such heterogeneous Web2.0 information but these approaches are limited in providing a common presentation of the results, which do not undergo any analysis phase. In the future work we aim to implement intelligent techniques able to analyse aggregated Web 2.0 content.

- **Linking** with **Open Data.** Exposing, sharing, and connecting pieces of data, information, and knowledge using URIs and RDF is crucial for the implementation of the Semantic Web. An additional goal of ours is to be able to enhance our techniques in order to be able to produce and analyse as much Linked Open Data as possible.
- Application in **other domains.** As discussed in section 2.3, there is a plethora of possibilities for the application of Collective Intelligence in other domains. In the future we intend to build Collective Intelligence applications suitable for use in the domains of news, finance, e-government, and local events, among others.

6. References

[1] Work, D, Blandin, S, Piccoli, B, Bayen, A (2010). A traffic model for velocity data assimilation. Applied Mathematics Research eXpress.

[2] Torres, LH (2007). Citizen sourcing in the public interest. Knowledge Management for Development Journal, 3(1): 134-145.

[3] iSpot, your place to share nature. URL: http://ispot.org.uk/.

[4] Mobnotes. URL: http://www.mobnotes.com/

[5] Dopplr. URL: http://www.dopplr.com/

[6] FixMyStreet. URL: http://www.fixmystreet.com/

[7] Malone, TW, Klein, M (2007). Harnessing Collective Intelligence to Address Global Climate Change. Innovations, 2(3):15-26.

[8] Kemp, C, Shafto, P, Berke, A, Tenenbaum, JB. Combining causal and similarity-based reasoning. Advances in Neural Information Processing Systems 19 , in press.

[9] Prelec, D (2004). A Bayesian Truth Serum for Subjective Data. Science, October 2004, 462-466.

[10] MIT Center for Collective Intelligence, Distributed collaboration project, URL: http://cci.mit.edu/research/collaboration.html

[11] Wang, C, Jing, F, Zhang, L, Zhang, H (2007). Content-based image annotation refinement. In CVPR.

[12] Liu, D, Hua, X, Wang, M, Zhang, H (2010). Retagging social images based on visual and semantic consistency. In WWW.

[13] Li, J, Wang, JZ (2003). Automatic linguistic indexing of pictures by a statistical modeling approach. TPAMI.

[14] Chang, EY, Goh, K, Sychay, G, Wu, G (2003). Cbsa: content-based soft annotation for multimodal image retrieval using bayes point machines. TCSVT.

[15] Jeon, J, Lavrenko, V, Manmatha, R (2003). Automatic image annotation and retrieval using cross-media relevance models. In SIGIR.

[16] Feng, S, Manmatha, R, Lavrenko, V (2004). Multiple Bernoulli relevance models for image and video annotation. In CVPR.

[17] Zhou, D, Bousquet, O, Lal, T, Weston, J, Scholkopf, B (2003). Learning with local and global consistency. In NIPS.

[18] Kennedy, LS, Naaman, M, Ahern, S, Nair, R, Rattenbury, T (2007). How Flickr helps us make sense of the world: context and content in

community-contributed media collections. ACM Multimedia, 631-640.

[19] Quack, T, Leibe, B, Van Gool, L (2008). World-scale mining of objects and events from community photo collections. In Proceedings of the 2008 international conference on Content-based image and video retrieval, 47-56.

[20] Girardin, F, Calabrese, F, Dal Fiore, F, Ratti, C, and Blat, J (2008). Digital footprinting: Uncovering tourists with user-generated content. IEEE Pervasive Computing, 7(4):36–43

[21] Hollenstein, L, Purves, RS (2009). Exploring place through user-generated content: using Flickr to describe city cores. Journal of Spatial Information Science.

[22] Blog of Collective Intelligence - http://www.community-intelligence.com/blogs/public/.

[23] Wikipedia, The Free Encyclopedia - http://www.wikipedia.org

[24] Bernardo A. Huberman. The social mind. In ean Pierre Changeux and Jean Chavaillon, editors, Origins of the Human Brain. Oxford: Clarendon Press, 1996

[25] Nanopoulos, A., Gabriel, H.H., Spiliopoulou, M.: Spectral clustering in social-tagging systems. In: WISE '09: Proceedings of the 10th International Conference on Web Information Systems Engineering, pp. 87–100. Springer-Verlag, Berlin, Heidelberg (2009).

[26] Specia, L., Motta, E.: Integrating folksonomies with the semantic web. In: ESWC '07: Proceedings of the 4th European conference on The SemanticWeb, pp. 624–639. Springer-Verlag, Berlin, Heidelberg (2007).

[27] Diederich, J., Iofciu, T.: Finding communities of practice from user profiles based on folksonomies. In: Proceedings of the 1st International Workshop on Building Technology Enhanced Learning solutions for Communities of Practice (TEL-CoPs'06) (2006).

[28] Schifanella, R., Barrat, A., Cattuto, C., Markines, B., Menczer, F.: Folks in folksonomies: social link prediction from shared metadata. In: WSDM '10: Proceedings of the third ACM international conference on Web search and data mining, pp. 271–280. ACM, New York, NY, USA (2010).

[29] Au Yeung, C.m., Gibbins, N., Shadbolt, N.: A study of user profile generation from folksonomies. In: SWKM (2008).

[30] Gemmell, J., Shepitsen, A., Mobasher, B., Burke, R.: Personalizing navigation in folksonomies using hierarchical tag clustering. In: DaWaK '08: Proceedings of the 10th international conference on Data Warehousing and Knowledge Discovery, pp. 196–205. Springer-Verlag, Berlin, Heidelberg (2008).

[31] Au Yeung, C.m., Gibbins, N., Shadbolt, N.: Contextualising tags in collaborative tagging systems. In: HT ’09: Proceedings of the 20th ACM conference on Hypertext and hypermedia, pp. 251–260. ACM, New York, NY, USA (2009)

[32] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors. In World Wide Web Conference, 2010.

[33] A. Signorini. Swine Flu monitoring using twitter. In http : //compepi.cs.uiowa.edu/ alessio/twitter − monitor − swine − flu/.

[34] Xin Jin, Andrew Gallagher, Liangliang Cao, Jiebo Luo, Jiawei Han . The Wisdom of Social Multimedia: Using Flickr For Prediction and Forecast, MM '10 Proceedings of the international conference on Multimedia, 25-29 October 2010, Firenze, Italy.

[35] Social Pixels: Genesis and Evaluation. Vivek K. Singh, Mingyan Gao, Ramesh Jain. MM '10 Proceedings of the international conference on Multimedia, 25-29 October 2010, Firenze, Italy.

[36] C. D. Corley and A. R. Mikler. A computational framework to study public health epidemiology. In International Joint Conferences on System Biology, Bioinformatics and Intelligent Computing (IJCBS09), Shanghai, China, August 2009.

[37] J. Ginsberg, M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant. Detecting influenza epidemics using search engine query data. Nature, 457:1012–1014, 2009

[38] G. Eysenbach. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. In AMIA 2006 Symposium Proceedings, pages 244–248, 2006.

[39] P. M. Polgreen, Y. Chen, D. M. Pennock, and F. D. Nelson. Using internet searches for influenza surveillance. Clinical Infectious Diseases (Supplement), pages 1443–1448, 2008.

[40] H. A. Johnson, M. M. Wagner, W. R. Hogan, W. Chapman, R. T. Olszewski, J. Dowling, and G. Barnas. Analysis of web access logs for surveillance of influenza. Stud Health Technol Inform., 107(Pt 2):1202´lC6, 2004.

[41] S. Papadopoulos, C. Zigkolis, S. Kapiris, Y. Kompatsiaris, A. Vakali. “City exploration by use of spatio-temporal analysis and clustering of user contributed photos” Demo paper accepted in ACM International Conference on Multimedia Retrieval (ICMR) 2011.

[42] Yannis Kalantidis, Giorgos Tolias, Yannis Avrithis, Marios Phinikettos, Evaggelos Spyrou, Phivos Mylonas, and Stefanos Kollias. 2011. VIRaL: Visual Image Retrieval and Localization. *Multimedia Tools Appl.* 51, 2 (January 2011), 555-592.

[43] WeKnowIt Deliverable D7.4.2 – "Consumer social group case study implementation", 25 Feb. 2011

[44] Max Braun, Ansgar Scherp, Steffen Staab (2010): Collaborative Creation of Semantic Points of Interest as Linked Data on the Mobile Phone, Extended Semantic Web Conf., Springer, 2010. Demo Session.

[45] Ovelgönne, M., Sonnenbichler, A. C., Geyer-Schulz, A., A Social Emergency Alert Service - A Location-Based Privacy-Aware Personal Safety Service, Fourth International Conference on Next Generation Mobile Applications, Services and Technologies, 26 – 29 July, Amman, Jordan

[46] Giannakidou, E., Kompatsiaris, I., Vakali, A.: Semsoc: Semantic, social and content-based clustering in multimedia collaborative tagging systems. In: ICSC, pp. 128–135 (2008)

[47] WeKnowIt Deliverable D7.3.2 – "Emergency response case study implementation", 31 Dec. 2010

[48] The CURIO ontology. http://staffwww.dcs.shef.ac.uk/people/G.Burel/ontologies/curio/

[49] N. Gilbert, K. Troitzsch (2005) Simulation and social science. Simulation for Social Scientists (2 ed.). Open University Press