A Resource Allocation Framework for Collective Intelligence System Engineering

Dimitrios J. Vergados
School of Electrical and Computer Engineering, National Technical University of Athens
Athens, 15780, Greece
+30 2107722538
djvergad@telecom.ntua.gr

Ioanna Lykourentzou
School of Electrical and Computer Engineering, National Technical University of Athens
Athens, 15780, Greece
+30 2107722538
ioanna@medialab.ntua.gr

Epaminondas Kapetanios
School of Computer Science, University of Westminster
London HA1 3TP, UK
+44 7849761529
e.kapetanios@wmin.ac.uk

ABSTRACT
In this paper, we present a framework for engineering collective intelligence systems that will be used by web communities. The proposed framework enables the development of community-driven, self-regulating CI systems, which adapt their functionality to the activity and goals of the web community. The above engineering methodology is applied on the design of a popular web system, namely Wikipedia, to illustrate the way that the functionality of the latter could be improved, in terms of better and more prompt article quality production. The preliminary evaluation results of this application, obtained through simulation modeling are promising.

Categories and Subject Descriptors
H. Information Systems, H.1 Models and principles

General Terms
Algorithms, Performance, Design, Human Factors, Standardization

Keywords
Collective intelligence, system engineering, model design, resource allocation

1. INTRODUCTION
One predominant type of digital eco-systems that has emerged in recent years is the one involving user collaboration, for instance the one developed through web 2.0 technologies. These technologies enable the collaboration and interlinking of web users around the globe towards a significant variety of purposes, which include socializing, e.g. through social networking sites and collaborating towards content development e.g. through Wikipedia, the popular web encyclopedia.

One of the reasons for the success of these ventures is the significant amount of participation and voluntarily efforts that they attract. This huge participation often leads to impressive results, with social networking sites taking up a large proportion of web users and the quality of a number of Wikipedia articles being compared to the quality of the same articles in Encyclopedia Britannica [5]. Nevertheless, the unsystematic manner that users participate in these systems has also raised a significant amount of concerns, mainly related to quality and timeliness. For instance a remarkable number of Wikipedia articles are still under-developed, or they reach high-quality levels in a slow, unpredictable pace.

An approach that could combine the advantages of high user participation and the merits of putting user skills into practice in a targeted manner could therefore tremendously boost the potential of web communities towards achieving their goals.

In this paper we propose a method to achieve the above, based on notions from collective intelligence [8], an idea which is listed as one of the six technologies likely to have the greatest impact worldwide [4].

In our approach, we view the issue of enhancing the results of web user collaboration as a resource allocation problem and seek, through algorithmic schemas, the optimal way of allocating the knowledge and skills of web users in order to better achieve their individual and community targets.

The constructed methodology aims at formally establishing the problem of engineering collective intelligence systems for human communities, at identifying the ground attributes that describe these systems and at modeling the parameters that will enable their effective engineering and optimization.

The rest of this paper is organized as follows: Section 2 presents the findings and limitations of current research literature. Section 3 presents the proposed CI system engineering methodology. Section 4, presents a potential application of this methodology on enhancing the functionality of a particular collaborative system, namely Wikipedia, in terms of quality and timeliness. Section 5 presents the evaluation results of the aforementioned application, which are obtained through simulation modeling. Finally, section 6 concludes with the main findings and future directions of the paper.

2. RELATED LITERATURE
A number of efforts approach the issue of enabling human communities to take advantage of their intellectual resources.

One of the most recent scientific innovations demonstrated towards this direction, is the idea of collaborative systems based on Web 2.0 technologies that permit knowledge gathering from a large number of users [18], based mostly on unsystematic contributions [9]. These collaborative systems are restricted by boundaries which are related mainly to their inability to automatically identify and, through proper coordination, make the most out of the knowledge that each individual has to offer. In that way they hold back user communities, limiting their role
to that of simple knowledge collection groups, preventing them from reaching their full potential in terms of content and solutions in a timely and reliable manner.

On the other hand, existing approaches regarding CI systems have the limitation of relying on the operation and stigmergic coordination of simplified unintelligent actors, such as simple software agents. As such they have been applied on more modest problems, which include computer network optimization [7], data mining [3], robotics [6, 13, 16] and machine learning [13, 15, 17], as well as on routing optimization problems [1].

The aforementioned CI approaches highly resemble swarm intelligence, i.e. the organizational format observed in natural communities, including insects and animals, where individual members perform very simple actions co-operating with one another to formulate results that cannot be achieved by individual members. However, these approaches cannot be directly transferred to human communities, since the behavior of the latter is governed by much more complex patterns. In order to elevate the intelligence of web communities, it is important to understand the fundamental differences between swarms and human user communities:

1. Firstly, the participants inside a swarm community have limited intellectual capabilities compared to the individuals of a human user community. Therefore although both formations – swarms and web users – exhibit the property of emergence, in the case of swarms this is the result of a stigmergic coordination among simple actors, whereas in the case of collaborating web users, emergence is the result of knowledge-oriented interactions among intelligent entities. Therefore, the development of a resource allocation schema for this different type of community needs to recognize and benefit from the above difference.

2. In addition, the actions performed by one swarm entity can also be performed by other entities in a similar manner. Nevertheless, in human user communities, individual actions are often unique and can be performed only by a specific individual, with the proper background, knowledge and skills. Therefore, research on CI systems should seek to identify and use the unique capabilities of each individual member of the community in an optimal manner.

3. Furthermore, swarm entities participate either by instinct, for instance in the case of biological swarms, or because they are programmed to do so, in the case of computer agents. It is essential to understand that, in the case of a CI system, web users are aware of the fact that they participate and thus they need to be motivated to do so.

4. Finally, in contrast to swarm entities, human behavior may often be unpredictable. Therefore it is essential to enable the machine intelligence, used to facilitate the resource allocation of the web community, to effectively function under non-deterministic circumstances.

Therefore, in order to truly fulfill their vision, collective intelligence systems need to be viewed from a totally different perspective, incorporate the unique traits of human intelligence into their functionality and coordinate intelligent actors, represented by human beings, into reaching higher-order decisions and results.

In addition, according to O'Reilly, collective intelligence applications need to fulfill the seven following requirements:

1. Task specific representations: Domain specific collective intelligence applications should support views of the task that are tailored to the particular domain.

2. Data is the key: Collective intelligence applications are data centric and should be designed to collect and share data among users.

3. Users add value: Users of collective intelligence applications know the most about the value of the information it contains. The application should provide mechanisms for them to add to, modify, or otherwise enhance the data to improve its usefulness.

4. Facilitate data aggregation: The ability to aggregate data adds value. Collective intelligence applications should be designed such that data aggregation occurs naturally through regular use.

5. Facilitate data access: The data in collective intelligence applications can have use beyond the boundaries of the application. Collective intelligence applications should offer Web services interfaces and other mechanisms to facilitate the re-use of data.

6. Facilitate access for all devices: The PC is no longer the only access device for internet applications. Collective intelligence applications need to be designed to integrate services across handheld devices, PCs, and internet servers.

7. The perpetual beta: Collective intelligence applications are ongoing services provided to its users thus new features should be added on a regular basis based on the changing needs of the user community.

Although the above principles provide significant insight regarding the direction towards which the design of CI systems should move, yet they do not provide a methodology for engineering CI systems in practice. Aiming at alleviating this issue, in this paper we propose a framework for the systematic design and development of CI systems. The proposed framework facilitates the development of CI systems which are able to adapt their functionality to the decisions and targets that the community will set, thus acting as facilitators that will enable it to self-organize, while at the same time they exhibit the emergence feature that has proven to benefit swarm communities.

3. A CI SYSTEM ENGINEERING FRAMEWORK

In this section the proposed framework for CI system engineering is presented. This framework is intended to be applicable on a wide variety of web communities and thus it
models CI systems from a higher level of abstraction. Designers wishing to apply it on a particular problem can further specialize it based on the requirements of the specific web community.

3.1 Generic CI system description
In general, a CI system comprises three main components, namely the human community, the machine intelligence, referred to also as the system engine, and the information that the system will at each time host (figure 1).

![Figure 1. CI system overview](image)

The community element provides the CI system with a number of capabilities that cannot be performed by machine intelligence, since they are unique human skills. These include the abilities of judgment, innovation, intellectual resources contribution and goal setting.

These community efforts lead to the information that the CI system hosts, including elements such as the individual actions performed, the decisions made and the solutions proposed.

Finally, the system engine is responsible for managing the intellectual and informational resources of the system. Intelligent algorithms are used, in this context, to gradually "learn" the capabilities of each community member, to identify the areas of the system where human interaction is required and to combine the above to allocate the system resources more efficiently. Therefore what the engine in essence does is coordinate user actions towards reaching their individual and community goals. Users are asked to perform localized tasks without the need of keeping pace with the complexity of all the activities that simultaneously take place. Yet, their aggregated actions, lead, through the task allocation schema managed by the engine, to the emergence of their intelligence as a community and to the best possible use of their potential.

It is important to note that another unique characteristic of the proposed CI system framework refers to community self-organization: that is while the engine actively guides individual community members towards performing specific tasks, it is itself at all times being driven by the implicit guidance and feedback of the community. In other words, all the decisions and the coordinating actions that the engine performs are dictated by the aggregated activities of the community as a whole. The engine therefore adapts to and becomes a reflection of the collective mind of the community, enabling it to self-organize and manage its own intellectual capabilities and potencies.

Summarizing, an efficiently engineered CI system will need to combine machine with the user intelligence, so as to formulate an unbreakably interconnected new type of intelligence system.

3.2 CI Modeling process
Engineering a CI system, which will combine machine and human intelligence in order to effectively synchronize user communities towards reaching higher-level results may be achieved in a variety of ways, depending on the specific requirements and goals of the community that will each time use the system. Therefore, the basic framework design may lead to different CI systems, which however all share a number of common characteristics and elements that need to be modeled [11]. These are briefly summarized in the following:

3.2.1 Important system attributes
**Community and Individual objectives**
Firstly, it is important to define the objectives that the community needs to reach through the use of the CI system, as well as the individual objectives that users will seek through its use. The clear definition of the above will facilitate the development of the subsequent resource allocation schemas.

**Set of possible individual actions**
The set of possible individual actions is another element that needs to be modeled. This set includes all the actions of the users towards the system, which can influence the given problem. They can include contributing ones intellectual skills, judging whether a result has been satisfactorily achieved, as well as other user-specific activities.

**System state**
The system state is another important attribute. This is defined as the minimal set of variables that may fully describe the important aspects of the system, including variables such as the problems set by the community, the solutions identified, the profiles and activity of the users, as well as the system resources.

3.2.2 Important functions
**Having modeled the important attributes of the CI system, one needs to also design a set of functions that will enable its efficient operation.**

**Expected community member action functions**
The definition of the functions that relate the current user activity to an accurate estimation of the expected future user activity is important, since it will enable the system to specify the capabilities of each individual and allocate them to the tasks where they will produce better results for both the individuals and the user community.

**Future system state functions**
Future system state functions are the ones that estimate the future state of the system, taking into account its current state, as well as the user activity that is expected to be performed in the specific time span.

**Objective functions**
The objective functions measure the level to which the community and individual objectives have been met through the allocation schema followed by the CI system so far.

3.2.3 CI system design and performance issues
Apart from the basic attributes and functions necessary to model the CI system, a number of other elements also need to be defined to ensure its effective functionality.

Resource allocation algorithms
One of the most basic elements of the CI modeling framework refers to the coordination and resource allocation algorithms. These algorithms define which actions will each specific community member be requested to perform at each specific state of the system, so that their community and individual objectives will be, to the best possible extent, maximized. Complementarily to maximizing the individual and community objectives, the basic modeling rationale of the resource allocation algorithms will need to fulfill the main requirement of the general CI system for community self-organization enabling.

Critical mass
An additional element that needs to be defined during the modeling process refers to critical mass, i.e. the minimum number of users necessary for the system to function effectively. Critical mass consists of those key users, whose behavior can significantly affect the collective behavior of the population, bringing it closer to its observed community intentions. This element is expected to be fully determined after the initial period of the CI system use.

Motivation
A final crucial element to be determined refers to user motivation. As stated above, in contrast to the fully predictable behavior of the simple agents performing in swarm intelligence systems, human users are aware of the fact that they participate in the CI system and thus their behavior is expected to be often unpredictable. However, motivation seems to play a significant role regarding the extent and quality of user participation and, to this end, it is crucial to model the specific CI system motivators. In general, there are three types of motivating factors to consider: tangible rewards [2], or more intrinsic motivators, such as social recognition [19] and the self-fulfillment element [12]. The type of motivator that will produce the best results is related to the community and individual objectives identified; yet it should be noted that whereas tangible rewards are expected to produce more prompt results, the incentives of intrinsic motivation seem to be more self-sustained [14].

4. CI SYSTEM CASE STUDY
Based on the framework presented above, this section presents a potential CI system case study. This case study refers to a user community that collaborates towards content creation. Since the most prominent example of such a community is Wikipedia, the case study presents how the application of the aforementioned framework can lead to a CI-enabled Wikipedia community that can produce better articles, in terms of quality and timeliness.

Wikipedia articles have often been criticized as having inadequate quality or as reaching satisfactory quality levels in a belated manner.

To address the above, we propose a CI system comprising the knowledge skills of the Wikipedia users and machine intelligence, which will allocate user contributions in such a way as to ensure that each inserted article reaches satisfactory quality levels within the time limits set by the community.

The main functionality of the system can be viewed as a self-regulating insert-[review-revise]-release spiral. As soon as a piece of information – a whole article or an article contribution - has been inserted into the system, a community review process takes place. During this process, the system selects the best possible group of community members to review the inserted piece of information and assess its quality. Then, consolidating these inputs the system either releases the information as reliable or, in case the information has not reached satisfactory quality levels, the system forwards it to selected members of the community for further enhancements. The processes of successive reviews and revisions continue, until the specific piece of information is deemed dependable and trustworthy, meeting the objectives set by the community (figure 2).

Figure 2. A CI-enabled Wikipedia system
This system satisfies both the requirement for collective intelligence emergence and the need for community self-organization.

As far as emergence is concerned, users act locally; they are requested to perform only a specific set of individual actions, which mainly include evaluating and contributing to selected content. They do not have to possess global knowledge over all the simultaneously performed activities, nor do they have to keep up with all the knowledge exchanges that take place inside the system. In parallel, machine intelligence in the form of the CI system engine retains an overview of all the activity taking place, and performs all the necessary computational actions that will leverage individual contributions into a higher-order community result. Such actions include assessing the quality of the information by accumulating individual evaluations, as well as allocating the reviewing and contribution tasks to the members of the community according to the constraints of time and quality that have been set.

As far as community self-coordination is concerned, it is crucial to emphasize that the engine makes its resource allocation decisions based on the implicit guidance of the community...
itself. For instance the decision whether a piece of information needs to be enhanced is directly derived from the aggregated ratings that it has received, whereas the users that the system asks to participate are selected based on characteristics such as their knowledge, an attribute also directly derived from the ratings that their past contributions have received by other users.

According to the approach proposed, individual users in the CI version of the Wikipedia system are organized in a self-regulating manner where each individual performs a specific set of actions, but their actions as a community are elevated, through the use of machine intelligence, to achieve their community and individual objectives.

In the following, we provide a summary of the values that the proposed CI system framework could receive for the specific Wikipedia enhancing system.

The community objective in this case is to produce articles of satisfactory quality and within specific time limits. The thresholds that define the quality levels and time constraints can be either pre-defined or they can be inferred based on user ratings. The individual user objectives in the case of Wikipedia seem to pre-exist and they mainly refer to contributing one’s efforts to the topics that one is mostly knowledgeable, as well as to the topics that interest them more. These objectives are also closely linked to the motivational factor, which in the specific CI system case is intrinsic and refers to self-fulfillment. As an additional motivator, the social recognition element could also be implemented, e.g. in the form of “top contributors” to further support user participation. The set of possible user actions of the system is relatively small and mainly consists of a) assessing the requested information and b) contributing their knowledge skills to enhance it. The system state can therefore be described with parameters referring to the hosted information, such as the article quality achieved so far, as well as to the users, such as workload, average time to submit a contribution, knowledge levels for each topic and so on.

The functions that determine the expected actions of the community members and the future state of the system can both be based on machine learning techniques. For instance, past user characteristics can be used to determine future user performance and activity. Machine learning can also be employed to match the specific characteristics of each inserted piece of information to the profile of each individual user. Through this matching procedure, the engine is enabled to accurately estimate the result that the contribution of each individual user would have if this individual were assigned with the specific piece of information. The objective functions are related to the extent to which the quality and time constraints have been met and they are used to enable system fine-tuning and enhancement. Finally, the resource allocation algorithms are the ones responsible for allocating tasks to users. Various allocation schemas can therefore be implemented ranging from the “most-appropriate-expert” policy, i.e. to select the most knowledgeable person for each task, to the “fairness-based” policy, i.e. to seek to allocate the task workload equally to the members of the community. Hybrid policies seeking to partially satisfy both the above constraints, as well as schemas that take into account the time factor can also be implemented.

Table 1 summarizes the main features of the proposed CI system.

**Table 1. Main framework features for the CI-enabled Wikipedia system**

<table>
<thead>
<tr>
<th>CI framework feature</th>
<th>Value</th>
</tr>
</thead>
</table>
| Community objective  | - Articles of satisfactory quality.  
|                      | - Articles released within specific time limits. |
| Individual objective | Contribute one’s knowledge |
| Set of possible individual actions | - Contribute  
|                      | - Review |
| System state         | - Information-related, e.g. current article quality  
|                      | - User-related, e.g. workload, time to submit a contribution, knowledge levels |
| Expected community member action functions | Machine learning, data-driven methods |
| Future system state functions | |
| Objective functions  | Functions measuring extent to which the quality and time criteria have been met. |
| Resource allocation algorithms | - Expert-based  
|                      | - Fairness-based  
|                      | - Hybrid schemas satisfying specific constraints (workload, time, expected contribution quality) |

5. EXPERIMENTAL RESULTS

Since the proposed CI system relies on the activity of a large number of web users, such a system is difficult to be evaluated prior to its actual deployment. However, a number of its advantages can be foreseen through simulation modeling. This assessment approach does not intend to achieve a fully accurate prediction of the system behavior, since it inevitably relies on more simplified models than those existing in real-world cases. Instead, it intends to illustrate the way that the CI-framework will improve the performance of the system in a qualitative manner and therefore to justify its performance prior to testing it in a large web user community.

The performance evaluation for the aforementioned system will be completed in two distinct phases: the design of the system model to be used during the simulation and the simulation itself.

5.1 Simulation model design

5.1.1 Real-world activity generator

The real-world activity generator is an abstraction of the expected behavior of the environment that interacts with the proposed system. It includes the functions and modeling elements necessary to generate the individual activity that will take place during the simulation of the system. It is crucial to note that the real-world activity generator is not known nor
taken into account by the simulated system, at any step of the process.

Thus, the most important features of the real-world generator are the following:

**Article:** Every article inside the system belongs to a specific knowledge domain and it is characterized by a certain quality level. The initial quality of each article is zero. Users may view an article or revise it, thus adding their own contribution to it. The revision of an article changes its quality. The level of improvement brought to an article through a user’s revision is determined taking into account the current quality of the article and the expertise of the user, i.e. the knowledge that the specific user has on the domain that the article belongs.

**User:** Each user is characterized by an expertise vector, which models their expertise on each one of the knowledge domains. Each value of the expertise vector is initialized independently for each user through a random variable.

### 5.1.2 System model

This feature models the important, to the evaluation, system elements. Thus, the system is modeled as follows:

Initially, the system comprises a large pool of users and a large base of articles. Users arrive to the system with a predetermined arrival rate and they randomly view an article. After viewing an article, a user may either revise it and therefore change its quality or simply leave without contributing. Since users are more likely to contribute to a subject that they know about, the user contribution probability to an article is modeled to be proportional to the user’s expertise in the domain of the article.

The above system is designed to resemble the current Wikipedia system, in which user skills are not actively taken into account by the system and thus users are not provided with specific requests or hints as to the articles that their contribution could be particularly helpful. This system design will serve as the benchmark that the proposed CI-enabled system will be compared to.

The CI-enabled Wikipedia system is modeled through the following allocation schema:

Whenever a user enters the system, it suggests him a specific article for revision, taking into account the user’s expertise. More specifically, the system identifies the domain in which the user has the most expertise. Then, for this domain, the system identifies the article that currently has the lowest quality and suggests it to the specific user.

Since whether a user will accept to contribute to a suggested article is expected to be affected only by the interest of the user to that specific article and not by the suggestion itself, users are modeled to respond to system requests with the same probability that they would respond in the benchmark system.

### 5.2 Simulation results

After designing the real-world generator and the CorpWiki system model, the simulation phase took place.

Prior to presenting the outcomes of the system evaluation, the simulation configuration features are presented. 10000 individual users were generated and 10000 articles were processed. The simulation lasted 10000 time units.

Firstly the quality of the articles produced through the use of the CI-enabled system was compared to the quality of the articles produced through the benchmark system. Figure 3 presents the results of this simulation scenario. As one may observe, the simulated user community managed to produce more qualitative articles through the use of the CI-enabled system, compared to the respective quality achieved through the use of the benchmark system.

![Figure 3. Comparison of the quality of the articles produced by the CI-enabled and the benchmark model](image)

Apart from quality, another important attribute of Wikipedia that needs to be considered refers to the time that an article reaches satisfactory quality, as well as to the number of human resources, measured through the number of revisions, needed to achieve the aforementioned quality levels.

As one may observe (figure 4), the quality achieved, during the same time span, by the CI-enabled system is higher than the respective quality achieved through the benchmark system. This finding is important, since Wikipedia articles often reach adequate quality levels in a slow manner and therefore a system that could speed up this process would be particularly helpful. Apart from time, one may also observe that the CI-enabled system has also saved the user community a number of valuable resources. That is, the number of revisions required for an article to reach a particular quality level is significantly higher for the benchmark system, compared to the resources used by the CI-enabled one.
The above results indicate that a CI-enable version of Wikipedia, which would combine user and machine intelligence and incorporate it into the system logic, could significantly help increase the produced article quality, better allocate user skills and reduce the time needed for the Wikipedia articles to reach to satisfactory levels of quality.

6. CONCLUSION
In this paper we propose a framework for the systematic engineering of collective intelligence systems designed for web user communities. The proposed framework is applied on the design of a popular system, namely Wikipedia, to depict how its functionality could be enhanced. The resulting CI system is evaluated through simulation modeling and certain preliminary outcomes are presented.

7. REFERENCES