Proceedings of the Thirtieth International Florida Artificial Intelligence Research Society Conference

# ViewpointS: When Social Ranking Meets the Semantic Web

**Philippe Lemoisson,[1] Guillaume Surroca,[2] Clément Jonquet,[2] Stefano A. Cerri[2]**
[1]UMR Territories, Environment, Remote Sensing and Spatial Information, CIRAD, Montpellier, France
[2]Laboratory of Informatics Robotics Micro-Electronics of Montpellier, France
philippe.lemoisson@cirad.fr;{surroca, jonquet, cerri}@lirmm.fr

## Abstract

Reconciling the ecosystem of semantic Web data with the ecosystem of social Web participation has been a major issue for the Web Science community. To answer this need, we propose an innovative approach called ViewpointS where the knowledge is topologically, rather than logically, explored and assessed. Both social contributions and linked data are represented by triples agent-resource-resource called "viewpoints". A "viewpoint" is the subjective declaration by an agent (human or artificial) of some semantic proximity between two resources. Knowledge resources and viewpoints form a bipartite graph called "knowledge graph". Information retrieval is processed on demand by choosing a user's "perspective" i.e., rules for quantifying and aggregating "viewpoints" which yield a "knowledge map". This map is equipped with a topology: the more viewpoints between two given resources, the shorter the distance; moreover, the distances between resources evolve along time according to new viewpoints, in the metaphor of synapses' strengths. Our hypothesis is that these dynamics actualize an adaptive, actionable collective knowledge.

We test our hypothesis with the MovieLens dataset by showing the ability of our formalism to unify the semantics issued from linked data e.g., movies' genres and the social Web e.g., users' ratings. Moreover, our results prove the relevance of the topological approach for assessing and comparing along the time the respective powers of 'genres' and 'ratings' for recommendation.

## Introduction

Since the last decade, reconciling the ecosystem of semantic Web data with the ecosystem of social Web participation has been a major issue for the Web Science community (Hendler et al. 2008). Each ecosystem has its own dynamics and challenges: the countless subjective contributions of the social Web yield folksonomies, but without centrally controlled coherence (Mika 2007; Mikroyannidis 2007) whereas semantically linking datasets whose models evolve independently is a never ending task (Karapiperis and Apostolou 2006). The frontier between them has nevertheless become porous: on one hand, the Web 2.0 contributors attempt to explore more and more Web data before posting comments and ranking contents; on the other hand the data analysis algorithms start to integrate social micro-contributions to improve their results. Bridging the gap between social ranking and the semantic Web is a big issue: it might be the key of a collective knowledge system able to bring in new levels of understanding in the sense coined by Gruber in (Gruber 2008).

The current paper is a proof of concept for an approach stepping forward in this direction called ViewpointS. After presenting our major sources of inspiration within the state of the art, we propose a formalism supporting the storage and exploitation of knowledge, and detail a topological assessment of information. Then we experiment all the novel aspects of our approach with the MovieLens dataset. Finally, we discuss the formalism and measurements and briefly expose directions for future work.

## Related work and inspirations

Facing the issue of bridging the gap between social ranking and the semantic Web, (Breslin, Passant, and Vrandečić 2011) distinguish two main options in the literature consisting respectively in "post-formalizing" and "pre-formalizing" the informal content of the social Web.

The ViewpointS approach takes a third option: to represent both social contributions and linked data by triples agent-resource-resource. The major consequence of this choice of a memory built upon connections made by agents is to let the collected knowledge evolve continuously along the interactions of the contributors. Another consequence is to ensure trustworthiness, well in line with the recent shift in the Web paradigm bringing the contributor back in the loop (O'Reilly 2015).

In ViewpointS, we started from a tripartite model agent-resource-tag generalizing social networks such as Flickr and proposed as building block the more abstract triple agent-resource-resource, in which agents may themselves be resources, expressing: "the belief of a resource/agent that two resources are close". This yields a topological, rather than logical, exploitation of the "wisdom of the crowd" in a manner similar to (Mika 2007; Markines et al. 2009; Specia and Motta 2007). Taking inspiration from the contexts of annotation, disambiguation, concept alignment and information retrieval (Harispe et al. 2014; Lee et al. 2008), we then grouped the triples in order to yield 'semantic similarity' or 'semantic proximity'. We firstly aggregate all the triples connecting two given *knowledge resources* into a higher level binary link called a *synapse*. In a second step, we made the hypothesis that any user's context could be translated into a set of quantification rules for the *viewpoints* called a *perspective*; this is in line with (Kim and Scerri 2008) which advice to evaluate ontologies on demand with respect to a specific context. Once a *perspective* is adopted, the initial heterogeneous semantics carried by the *viewpoints* are transformed into a *knowledge map* equipped with a distance. As a consequence, the agents can use the proximities resulting from the existing *viewpoints* when browsing *knowledge maps* and reversely update the knowledge through new *viewpoints* expressing their feedback e.g., "I believe this agent matches/does-not-match the topic of my query" or "I deny concept-a/ instance-1 and concept-b/ instance-2 are connected through relation-r". Along these exploitation/feedback cycles, the shared knowledge is continuously elicited against the beliefs of the agents in a selection process supported by the evolving strength of *synapses*. Here comes the metaphor of the brain. According to the Theory of Neuronal Group Selection (Edelman and Tononi 2000), which recently reappeared in the front scene under the name of connectome (Seung 2012), knowledge results from a process of continual re-categorization. In our approach, the dynamics of synapses aim at yielding an evolving topology where knowledge is constantly reorganized[1].

The next section firstly presents the formalism supporting the ViewpointS "Knowledge graph", then goes through some aspects exploring its topology in order to assess collective knowledge on top of heterogeneous semantics.

## A topological, unified view on heterogeneous semantics

In the ViewpointS approach, all the resources (identified by a URI) contributing to knowledge are grouped in a single class: *Knowledge resource*. Figure 1 illustrates five mutually exclusive subclasses of knowledge resources covering most of the practical cases:

[1] This point requires a real-life scenario and cannot be demonstrated in this paper; setting such a scenario is our next objective.

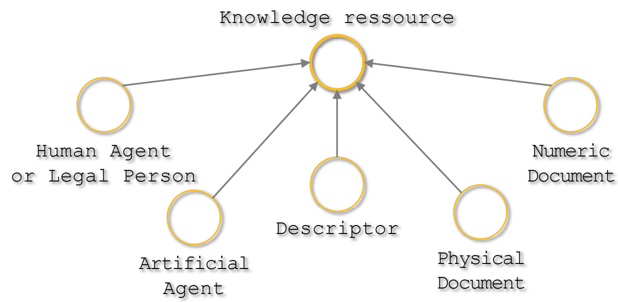


*Figure 1: Knowledge resources*

- Human Agent or Legal person: entity "performing acts and undertaking obligations" [32] such as humans or organizations, by emitting connections between resources.
- Artificial Agent: numeric entity emitting connections between resources.
- Physical document: document of the real world such as a book.
- Numeric document: numeric entity such as a Web page.
- Descriptor: meaningful linguistic expression, playing the role of tag.

*Knowledge resources* participate to connections called *viewpoints*: each *viewpoint* is a subjective connection established by an agent (Human or Artificial) between two knowledge resources; each *viewpoint* implicitly brings in the specific semantics of its emitter. However, these semantics do not need to be shared by the agents that will further exploit the knowledge, as it will be explained when describing the "knowledge maps". The formalization of these heterogeneous semantics is illustrated in Figure 2:

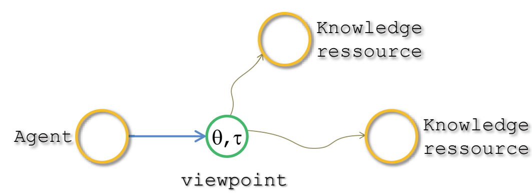


*Figure 2: A "viewpoint". The straight arrow gives the provenance; 'θ' gives the semantics, 'τ' gives the time stamp*

The viewpoint ($a_1$, {$r_2$, $r_3$}, θ, τ) stands for: the agent $a_1$ believes at time τ that $r_2$ and $r_3$ are related according to the semantics carried by θ. For instance:

The viewpoint (AA-MovieLens, {movieX, genreY}, mo:genre, $\tau_1$) stands for: the artificial agent 'AAMovieLens' declares at $\tau_1$ that the physical document 'movieX' matches the descriptor 'genreY'; in this example, θ=mo:genre is a standard unambiguous semantic Web property issued the MovieLens linked data.

The viewpoint (userX, {movieY, ***}, vp:rating, $\tau_2$) stands for: the legal person 'userX' declares at $\tau_2$ that the physical document 'movieY' matches the descriptor '*** =

of average interest'; in that example, θ=vp:rating is a transcription of the MovieLens linked data.

The *viewpoints*'`types` are grouped into 7 `meta-types` `:`{author, like, similar, onto-match, social-match, algo-match, preview}.

## The knowledge graph

*Knowledge resources* and *viewpoints* together form a bipartite *Knowledge graph*[2] denoted `KG`, where the knowledge is persistently stored. By contrast, retrieving information happens in transient maps computed on demand.

## The knowledge maps

A preliminary step for retrieving knowledge consists in reifying the querying context by use of a *perspective* `U` i.e., a set of quantification rules applied to the *viewpoints*. It may be default rules adopted by a group of users in a recurrent context, or specific rules filtering `KG` according to preferences such as: ignoring the *viewpoints* anterior to a given date, privileging the *viewpoints* emitted by some agents or privileging *viewpoints* of a given `meta-type`.

Given a `KG`, computing the *knowledge map* associated to a context `U` involves three steps: i) grouping all the *viewpoints* connecting any given pair of *knowledge resources* into higher level link called *synapses* and then ii) choosing the rules for valuating the *synapses* i.e., setting the *perspective* `U` as a map-reduce process. This is illustrated in Figure 3.

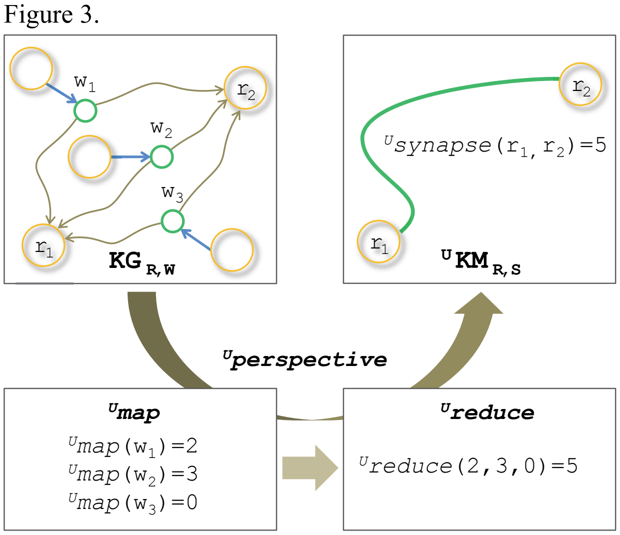


*Figure 3: Building a knowledge map $^{U}KM$*

We call *knowledge map* and denote $^{U}$`KM` the undirected labelled graph interpreting `KG` through the perspective `U`. Depending on the *perspective* adopted, one single `KG` may therefore be interpreted into several distinct $^{U}$`KM`, the topology of which can be exploited with standard graph algorithms in order to exhibit knowledge. In practice, $^{U}$`KM` is never built exhaustively! Instead, the *synapses* are computed on demand along a Dijkstra-inspired exploration bounded by a parameter `'m'`. We denote $^{U,m}$ Ψ-`neighbours(r)` the neighborhood of a target `'r'` resulting from an exploration bounded by `'m'` in the perspective `U`. This can be compared to the semantic similarity of (Zhu and Iglesias 2017); however using the perspective `U` only requires knowledge about `meta-types` i.e., the user does not need any preliminary knowledge about the ontologies source of the *viewpoints* of `meta-type` "onto-match".

Two things should be noted: i) the bigger the *synapse* the shorter the distance and ii) as long as `'m'` is small compared to the size of `KG`, the exploration of new branches quickly stops; therefore the worst-case complexity[3] is never reached in practical cases. The practical complexity does not depend on the size of `KG` but on its local density.

## The topological assessment of the knowledge

(Adomavicius and Tuzhilin 2005; Yamaba, Tanoue, and Takatsuka 2013) write that powerful recommender systems will exploit more and more the underlying topologies. Taking inspiration from this, our objective is to provide a means to characterize the underlying topology in `KG` in terms of information. For example, if when browsing through a given movie dataset the uncertainty about genres is reduced by its topology, we may say according to (Klir 2005) that the dataset embeds information about 'genres'.

Let us consider in `KG` a collection 'O' of `objects` (e.g., 'movies') and a collection 'D' of `descriptors` (e.g., 'genres'), we measure whether close elements of 'O' have similar elements of 'D' in their respective neighborhoods. We call "local homogeneity" (of O with respect to D) the probability to find similar elements of 'D' in the neighborhoods of close elements of 'O'. Let |D| be the cardinal of D, in order to compute the "local homogeneity" we use a |D|-dimensional vector space.

> Let `U` be a perspective, let `'m'` be the parameter for computing neighborhoods, let $O=\{o_i\}_{i\in|O|}$ and $D=\{d_j\}_{j\in|D|}$ be two collections of *knowledge resources*, we denote $D(o_j)$ the vector of dimension $|D|$ $\{d^j(o_i)\}_{j\in|D|}$ such that $d^j(o_i)$=number of occurrences of $'d_j'$ in $^{U,m}$ Ψ-`neighbours`$(o_i)$.
> $^{U,m,D}$ `localHomogeneity(O)` is the average value of cosine similarity $(D(o_1), D(o_2))$ computed upon all the pairs $(o_1, o_2)$ verifying $^{U}\Psi(o_1, o_2) \leq m$.
> This will be denoted: $^{U,m,D}$ `lH(O)`.

[2] Although our approach has not yet addressed large datasets such as DBpedia, it has been thought to do so.

[3] The worst case complexity of the algorithm is $O(|W|^2|R|^2)$.

The value of local homogeneity varies between '0' and '1'. A local homogeneity of '0' corresponds to the absence of information expressed by D with respect to O. A local homogeneity of '1' corresponds to the maximal information expressed by D with respect to O. Since the computation of "local homogeneity" only comes into play for remotely assessing the evolution of collective knowledge, we do not worry about the complexity[4] resulting from the size of |D|.

## The MovieLens Experiment

To prove the concept of ViewpointS, we take a Web dataset where explicit knowledge expressed by linked data is mixed with implicit knowledge issued from social contributions: MovieLens[5]. The complete dataset consists of 100.000 movies and 1.000.000 ratings which have been collected by the GroupLens Research Project at the University of Minnesota. Their data have played the role of experimental matter for many authors (Peralta 2007; Harpale and Yang 2008; Jung 2012). In the following, we take an extract of the original dataset: 1682 movies rated by 943 commenters providing 5.000 ratings. We aim at illustrating:

a) how a Web dataset involving linked data from the semantic Web together with individual opinions issued from social ranking can be transcribed into a unique *knowledge graph*;
b) how the implicit knowledge associated to 'ratings' and the explicit knowledge issued from linked data can be both (and yet distinctly) topologically assessed.

### The ViewpointS model for MovieLens

To transcribe the explicit information embedded in MovieLens into ViewpointS, we firstly create *knowledge resources*:

- `MovieId` stands for movies as *Physical documents*
- `UserId` stands for users as *Legal persons*
- `AA-Movielens` is the *Artificial agent* corresponding to the MovieLens Ontology
- `Age`, `genreId`, `gender`, `Occupation`, `Year` are considered as *Descriptors*

We then consider that `AA-MovieLens` is the emitter of *viewpoints* with `meta-type` "onto-match[6]":

- the *viewpoint* `(AA-MovieLens, {movieX, genreY}, mo:genre,τ)` stands for: the `artificial agent 'AA-MovieLens'` connects the `object 'movieX'` to the `descriptor 'genreY'`

We then consider that each `userId` emits self-describing viewpoints with `meta-type` "onto-match":

- the *viewpoint* `(userId, {userId, occupationX}, cv:jobType,τ)`

We finaly consider that each `userId` rate movies through viewpoints with `meta-type` "social-match":

- the *viewpoint* `(userX, {movieY, ***}, vp:rating,τ)` stands for: the `legal person 'userId'` considers the `object 'movieY'` as `'of average interest'`

### Protocol and measurements

We initialize the *knowledge graph* by transcribing the movies descriptions and users profiles as explained upper. Each movie is linked to one or more of the 18 'genres' by *viewpoints* of type `mo:genre` and to one 'year' by a *viewpoint* of type `mo:relaseDate`. Each user is linked to one 'class of age' by a *viewpoint* of type `foaf:member`, one 'gender' by a *viewpoint* of type `foaf:gender`, and one 'occupation' by a *viewpoint* of type `cv:jobType`. All these *viewpoints* are time-stamped $\tau_0$; this is called $\mathtt{cycle}_0$. We then arbitrary split the ratings into 5 subsets of 1000 ratings each (called $\mathtt{cycle}_1$ to $\mathtt{cycle}_5$), using *viewpoints* of type `vp:rating` time-stamped $\tau_i$, where 'i' is the cycle number.

We consider 3 perspectives:

- $U_1$ reflects priority given to explicit knowledge, i.e., *viewpoints* of `meta-type` "social-match" are weighted 1 whereas *viewpoints* of `meta-type` "onto-match" are weighted 3. In this perspective, the distance between a 'movie' and its 'genre' is 1/3, so that consequence two movies of same 'genre' are at distance <= 2/3; we therefore expect 'm=1' to be relevant for measuring the local homogeneity in 'genres' through this perspective.
- $U_2$ reflects balanced importance of explicit versus implicit knowledge, i.e., all *viewpoints* are weighted 1.
- $U_3$ reflects priority given to implicit knowledge, i.e., *viewpoints* of `meta-type` "social-match" are weighted 3 whereas *viewpoints* of `meta-type` "onto-match" are weighted 1.

In the *perspectives* $U_2$ and $U_3$, *viewpoints* of type `mo:genre` are valuated '1', so that two movies connected to the same 'genre are at distance' <= 2; we therefore expect 'm=2' to be relevant for measuring the local homogeneity in 'genres' through these two perspectives.

Besides, the multiplicity of ratings by different users in the dataset will yield strong synapses and in consequence short distances. As a consequence, both 'm=1' and 'm=2'

[4] It might happen that we later need to undergo dimensional reduction in a way similar to the "AffectiveSpace" of (Cambria et al. 2015).
[5] http://datahub.io/dataset/movielens
[6] We use the semantic Web properties in the transcription process as often as possible, e.g., mo:genre, where the prefix 'mo:' refers to the Movie Ontology (Bouza 2010) which unambiguously defines its resources and properties. Doing so, we enable ViewpointS to embrace the semantic Web standards and import/export RDF representations.

are expected suitable for measuring the local homogeneity in ‘ratings’, whatever the perspective.

The experiment consists in going through $\mathtt{cycle}_0$ to $\mathtt{cycle}_5$ in 3 successive runs and measuring “local homogeneities” at the end of each cycle of each run.

- $\mathrm{RUN}_{1,1}$ corresponds to the perspective $\mathrm{U}_1$ with m=1
- $\mathrm{RUN}_{2,2}$ corresponds to the perspective $\mathrm{U}_2$ with m=2
- $\mathrm{RUN}_{3,2}$ corresponds to the perspective $\mathrm{U}_3$ with m=2

## Denotations

- M is the population of 1682 movies
- Genres is the collection of 18 genres
- Ratings is the collection {*, **, ***, ****, *****}
- ${}^{\mathrm{Ui,m,}\mathbf{Genres}}\mathrm{lH}(\mathrm{M})$ is the local homogeneity of movies with respect to genres
- ${}^{\mathrm{Ui,m,Ratings}}\mathrm{lH}(\mathrm{M})$ is the local homogeneity of movies with respect to ratings

## Results and interpretation

The 3 runs corresponding to the three perspectives associated with the suitable ‘m’ parameters (namely $\mathrm{RUN}_{1,1}$, $\mathrm{RUN}_{2,2}$ and $\mathrm{RUN}_{3,2}$) are presented in Figure 4.

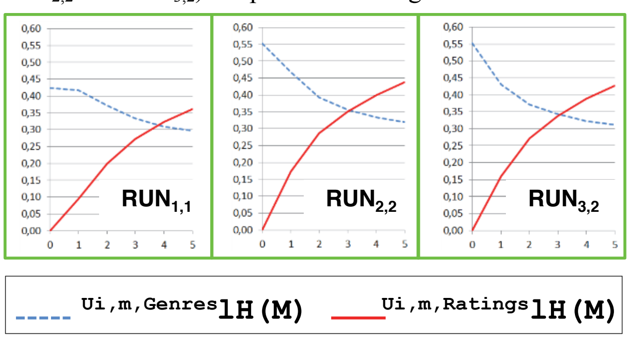


*Figure 4: Assessment of explicit knowledge (genres) and implicit knowledge (ratings) along the interaction cycles*

We firstly observe that ${}^{\mathrm{Ui,m,}\mathbf{Genres}}\mathrm{lH}(\mathrm{M})$ always takes significant values (between 0.30 and 0.55), whatever the cycle and the perspective. We interpret it as “genres provide information about movies, despite the noise due to other informational dimensions”.

We then observe that ${}^{\mathrm{Ui,m,}\mathbf{Ratings}}\mathrm{lH}(\mathrm{M})$ always increases along the cycles (between 0.00 and 0.45), whatever the cycle and the perspective. We interpret it as “ratings provide information about movies, roughly in proportion with their number, despite the noise due to other informational dimensions”.

The major output of the experiment is that the value of ${}^{\mathrm{Ui,m,}\mathbf{Genres}}\mathrm{lH}(\mathrm{M})$ decreases while ${}^{\mathrm{Ui,m,Ratings}}\mathrm{lH}(\mathrm{M})$ increases along the cycles, whatever the run. This illustrates concurrence and complementarity between the topological assessments respectively associated to ‘genres’ and to ‘ratings’. We interpret it as “genres and ratings are independent informational dimensions”.

Moreover, the intersection of the curves corresponding to these two dimensions occurs after $\mathrm{cycle}_3$ i.e., after 3000 *viewpoints* ‘rating’, to be compared with 2893 *viewpoints* ‘genre’. We interpret it as: “ratings and genres are equally powerful informational dimensions”.

Finally, the curves intersect at a point quite independent from the perspective. This reinforces the hypothesis that “local homogeneity” proves information within the hidden topology of KG, whatever the map. A close look shows that the highest local homogeneity at intersection point appears in the perspective $\mathrm{U}_2$: 0.32 for $\mathrm{RUN}_{1,1}$, 0.35 for $\mathrm{RUN}_{2,2}$, 0.34 for $\mathrm{RUN}_{3,2}$. We interpret it as “the balanced perspective $\mathrm{U}_2$ is the best suited for observing simultaneously genres and ratings”.

## Discussion

The MovieLens dataset has fulfilled our demonstration objectives:

a) we have transcribed all the information stored in the MovieLens dataset (social ranking and explicit semantics) into a unique *knowledge graph*.
b) we have topologically assessed the progressive dissemination of the implicit knowledge associated to ‘ratings’, in a knowledge graph initially embedding only the explicit knowledge of linked data.

Let us discuss these two points more in detail. Our formalism seems suitable for capturing both explicit knowledge delivered by linked data and implicit knowledge available in the Social Web. We cannot pretend however to capture all the richness of the semantic Web: for instance, the *viewpoints* fail in expressing the conceptual ‘parent-child’ relationship or ‘composition-component’ relationships, usually known as ‘is a’ and ‘is part of’ respectively. We capture only flat descriptions, not the conceptual verticality of ontologies. This built-in limitation rules out the possibility of logical assessment, but allows the topological assessment of proximity in a context of heterogeneous semantics.

Moreover our unified characterization of the collective knowledge seems well suited for observing the Web dynamics: we have provided topological arguments for assessing that “genres and ratings are complementary dimensions competing for providing information about movies recommendation, with comparable power”.

Another important aspect of the approach is the disconnection between the storage of the knowledge events (the *viewpoints*), and their delayed interpretations through *knowledge maps* responding to distinct *perspectives*. The *viewpoints* are purely qualitative; interpreting them in terms of quantities entirely depends on the perspective taken by the final user. *Perspectives* can be seen as a kind of global social ranking occurring “at exploration time”. This might be a response to the usual biases in social networks resulting from self-promotion or dishonest recommendation, since untrusted viewpoints emitters can easily be discarded when tuning the perspective.

## Future work

We are currently developing an API offering intuitive input, easy browsing of the knowledge and one-click feedback. The next step in our agenda is to prove the concept in real life scenarios, i.e., to invite users to elicit knowledge collectively by using the ViewpointS approach through the API mentioned above. Two use cases have been planned, both oriented towards cross-disciplinary discoveries: one in the biomedical domain, in the context of the SIFR project, the other in the agronomic domain will be hosted by Cirad.

## Acknowledgements

This work was supported in part by the French National Research Agency under JCJC program, grant ANR-12-JS02-01001, as well as by University of Montpellier, CNRS and the CIRAD.

## References

Adomavicius, G.; and Tuzhilin, A. 2005. Toward the next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE Transactions on Knowledge and Data Engineering* 17 (6). IEEE: 734–749.

Bouza, A. 2010. MO - the Movie Ontology | Leverage Your Movie Information, http://www.movieontology.org/ (accessed on february 20th 2017).

Breslin, J. G.; Passant A.; and D Vrandečić D. 2011. Social Semantic Web. In *Handbook of Semantic Web Technologies*, 467–506. Springer Berlin Heidelberg.

Cambria, E.; Fu J., Bisio, F.; and Poria, P. 2015. AffectiveSpace 2: Enabling Affective Intuition for Concept-Level Sentiment Analysis. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI'15). *AAAI*, 508–514.

Edelman, G. M., and Tononi, G. *A Universe of Consciousness: How Matter Becomes Imagination*. 2000. Basic Books - 274 pages

Gruber, T. 2008. Collective Knowledge Systems: Where the Social Web Meets the Semantic Web. *Web Semantics: Science, Services and Agents on the World Wide Web* 6 (1): 4–13. doi:10.1016/j.websem.2007.11.011.

Harispe, S.; Ranwez, S.; Janaqi, S.; and Montmain, J. 2014. The Semantic Measures Library and Toolkit: Fast Computation of Semantic Similarity and Relatedness Using Biomedical Ontologies. *Bioinformatics (Oxford, England)* 30 (5): 740–742.

Harpale, A. S.; and Yiming, Y. 2008. Personalized Active Learning for Collaborative Filtering. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '08*, 91. New York, USA: ACM Press.

Hendler, J.; Shadbolt, N.; Hall, W.; Berners-Lee, T.; and Weitzner, D. 2008. Web Science: An Interdisciplinary Approach to Understanding the World Wide Web. *Communications of the ACM* 51 (7): 60–69.

Jung, J. J. 2012. Attribute Selection-Based Recommendation Framework for Short-Head User Group: An Empirical Study by MovieLens and IMDB. *Expert Systems with Applications* 39 (4): 4049–4054.

Karapiperis, S; and Apostolou, D. 2006. Consensus Building in Collaborative Ontology Engineering Processes. *Journal of Universal Knowledge Management*, 199–216.

Kim, H.L.; and Scerri, S. 2008. The State of the Art in Tag Ontologies: A Semantic Model for Tagging and Folksonomies.

Klir, G. J. 2005. *Uncertainty and Information: Foundations of Generalized Information Theory*. 2005. Wiley-IEEE Press

Lee, W-N.; Shah N., Sundlass, K.; and Musen, M. 2008. Comparison of Ontology-Based Semantic-Similarity Measures. *AMIA, Annual Symposium 2008*, January, 384–388.

Markines, B.; Cattuto, C.; Menczer, F.; Benz, D.; Hotho, A.; and Stumme, G. 2009. Evaluating Similarity Measures for Emergent Semantics of Social Tagging. In *18th International Conference on World Wide Web - WWW '09*, 641–50. Madrid, Espagne: ACM Press.

Mika, P. 2007. Ontologies Are Us: A Unified Model of Social Networks and Semantics. *Web Semantics: Science, Services and Agents on the World Wide Web* 5 (1): 5–15.

Mikroyannidis, A. 2007. Toward a Social Semantic Web. *Computer*. 40(11): 113-115.

O'Reilly, T. 2015. What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. http://papers.ssrn.com/abstract=1008839 (accessed on november 2016).

Peralta, V. 2007. Extraction and Integration of Movielens and Imdb Data. Laboratoire Prisme, Université de Versailles, Versailles, France.

Seung, S. 2012. *Connectome: How the Brain's Wiring Makes Us Who We Are*. Houghton Mifflin Harcourt.

Specia, L.; and Motta, E. 2007. Integrating Folksonomies with the Semantic Web. In ESWC '07 Proceedings of the 4th European conference on The Semantic Web: Research and Applications. Pages 624-639. Innsbruck, Austria. Springer-Verlag.

Yamaba, H.; Tanoue, T.; and Takatsuka, K. 2013. On a Serendipity-Oriented Recommender System Based on Folksonomy. *Artificial Life and Robotics* 18(1-2): 89–94.

Zhu, G.; and Iglesias, C. A. 2017. Computing Semantic Similarity of Concepts in Knowledge Graphs. *IEEE Transactions on Knowledge and Data Engineering* 29 (1): 72–85.