

Set Semantic Similarity for Image Prosthetic Knowledge Exchange

Valentina Franzoni[1,2(✉)], Yuanxi Li[3], and Alfredo Milani[1]

[1] Department of Mathematics and Computer Science, University of Perugia, via Vanvitelli, 1, 06123 Perugia, Italy
valentina.franzoni@dmi.unipg.it
[2] Department of Computer, Control, and Management Engineering "Antonio Ruberti", Sapienza University of Rome, via Ariosto, 25, 00185 Rome, Italy
[3] Department of Computer Science, Hong Kong Baptist University, Kowloon, Hong Kong, China
csxyli@comp.hkbu.edu.hk

Abstract. Concept information can be expressed by text, images or general objects which semantic meaning is clear to a human in a specific cultural context. For a computer, when available, text with its semantics (e.g., metadata, comments, captions) can convey more precise meaning than images or general objects with low-level features (e.g., color distribution, shapes, sound peaks) to extract the concept underlying the object. Among semantic measures, web-based proximity measures e.g., confidence, PMING, NGD, Jaccard, Dice, are particularly useful for concept evaluation, exploiting statistical data provided by search engines on terms and expressions provided in texts associated with the object.

Where Artificial Intelligence can be a support for impaired individuals, e.g., having disabilities related to vision and hearing, understanding the concept underlying an object can be critical for an intelligent artificial assistant. In this work we propose to use the set semantic distance, already used in literature for semantic similarity measurement of web objects, as a tool for artificial assistants to support knowledge extraction; in other words, as prosthetic knowledge.

Keywords: Information retrieval · Artificial Intelligence · Assistants · Semantic proximity · Group distance

1 Introduction

With the persistent diffusion of digital social interactions, the investigation of the information conventional by human-machine interactions became a critical research problem, as an increasing quantity of economic, personal and physical connections is digitally intermediated, stored and tracked, and can be used to predict and influence the way people or entities interact, vote, work, and behave [1, 2]. The snowballing amount of information daily produced using mobile devices, e.g., images, videos, music, with explicit (e.g., tags, hashtags, and keywords) or implicit (e.g., captions or metadata) content, can augment the information value of the document. Such informative content

© Springer Nature Switzerland AG 2019
S. Misra et al. (Eds.): ICCSA 2019, LNCS 11624, pp. 513–525, 2019.
https://doi.org/10.1007/978-3-030-24311-1_37

can be called Prosthetic Knowledge (PK), as the evidence that a user does not know can be accessed using technology.

When such information is not retrievable with high usability, an average user is informatively-impaired, even if without having physical or psychic disabilities. On the other hand, when the information is available in proper media, even a physical or psychic impaired user can become fully able to benefit from it.

In this environment, PK is the latent knowledge underlying augmented information in augmented documents, generated and eventually shared by users using technologies in high-tech devices and that, stored in structure or semi-structured form, can be used by intelligent assistants to enhance the information retrieval capabilities of the device, thus of the user [3–5].

In the following sections, the use of a set semantic similarity scheme is proposed for prosthetic knowledge. In Sect. 2, Concept-based Image Retrieval is presented with its pros and cons. In Sect. 3, related works on set semantic image similarity are summarized. In Sect. 4, experiments on the efficacy of the set proximity scheme for image similarity calculation is proposed as an example of the application of prosthetic knowledge. In Sect. 5, conclusions are drawn.

2 Concept-Based and Content-Based Image Retrieval

In *Content-based* Image Retrieval, proximity measures of low-level features [6] guide the process of retrieval. Content-based Image Retrieval aims to involve the proximity from a perceptual point of view, and deep-semantic relations among images cannot be discovered, because they require to integrate Content-based similarity with the broader notion of *Concept-based* Image similarity, [7] more useful for retrieval aims, as shown in Fig. 1.

Opening with the reflexion that the textual information related to images, such as captions, labels, tags, and comments, convey more concept-related than content-based information, we can concentrate on the textual context e.g., metadata and social networks links among objects, media and users themselves.

The features of big networks of concepts, built collaboratively by a collaboration of users, provide numerous advantages, among which:

1. *Information sharing and explanation direction*: context knowledge is annotated and filtered manually, with the unambiguous purpose of knowledge sharing, e.g., the hyperlink structure of Wikipedia.
2. *Open end and dynamic update*: collaborative networks of concepts are nonstop updated by a crowd of users, enhancing the temporal and qualitative value of the information, where new knowledge is timely added to the network as soon as such information arises. Semantic ontologies, such as WordNet, [8] tend on the other hand to be more stationary and to reflect the cultural or personal biases of their authors, due to the smaller number of contributors, even if experts.
3. *Heterogeneousness*: collaborative and social networks contain information about several different objects, users and relations as sources to deduce semantic

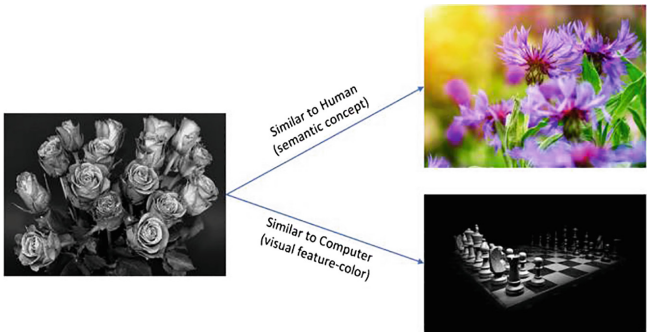

Fig. 1. Different similarity connotations between human and computer.

knowledge; the objects/documents in the network are included in a relevant semantic context.

Such features are useful for some applications, including:

- ***Query expansion***, where the notion of the original term in the query can be extended with a set of suitable keywords in order to retrieve related objects. Our method exploits the textual information of entities, relying on outward web-based knowledge bases. Measuring the proximity of objects by scheming the similarity of their related annotations can be productive using similarity/distance measures. Such way exploits basic web-based semantic proximity between two objects (i.e., among their textual annotations), and our proposed method delivers a new class of measures for group/set distance and context extraction, composing concept similarity measures.
- ***Prediction*** of annotation of images using prosthetic knowledge, [9, 10] and the suggestion of explicit annotations with image sub-structures, to predict their similarity.

3 Semantic Similarity Measures

We propose to use for prosthetic knowledge a set distance built on web-based proximity measures. Such a distance is a combination of selected values obtained by the calculation of several web-based proximity measures, some of which presented in detail in the following sub-paragraphs: first, related works are presented, then our set distance is exposed. Statistic-based measures can be straightforwardly calculated from search engine statistical results, while ontology-based measures require to navigate static ontologies (e.g., Wordnet) or dynamic ontologies (e.g., Wikipedia). It is worth to note that in both cases of statistics on search engines and dynamic ontologies, the measures can reflect the dynamic information underlying the knowledge of the users uploading data.

3.1 Ontology-Based Semantic Similarity

A. WordNet Distance [8, 11, 12]

The *WordNet Distance* (WD) defines the similarity between two concepts x and y in WordNet, exploiting the information about the concepts that subsume the given concepts in the ontology:

$$D_{wnet}(x, y) = \frac{2 \times \log P(lso(x, y))}{\log P(x) + \log P(y)} \tag{1}$$

where *lso(x, y)* is the lowest super-ordinate (i.e., most specific common subsumer) of concepts x and y, i.e., the closest common ancestor in the ontology, and *P(z)* is the probability of hitting a *synset* (i.e., synonyms set, the fundamental entity in WordNet) z, using a Shannon measure of the information content [13].

B. Wikipedia Distances

In Wikipedia, the hyperlinks among articles build a semantic explanation network among the pages.

The *WikiRelate* model [14] has the achievement to be the first model to compute semantic proximity measures using Wikipedia relationships. This approach uses techniques previously applied to WordNet and modifies them for Wikipedia's links category hierarchy.

Second, among the first models to use Wikipedia, the *Wikipedia Link Vector Model* [15, 16] uses the hyperlink structure of Wikipedia to provide a structured knowledge base without considering the categories anymore. The probability of the D_{wiki} distance in the Wikipedia Link Vector Model is defined as the fraction of the number of links to a target page on the total number of pages. Thus, if t is the total number of articles within Wikipedia, then the weighted value for the link $x \rightarrow y$ is:

$$D_{wiki}(x \rightarrow y) \; = \; |x \rightarrow y| \; \times \; \log(\sum_{i=1}^{t} \frac{t}{|i \rightarrow y|}) \tag{2}$$

where x and y denote the two articles about terms x and y.

To be mentioned, the set of the *Heuristic Semantic Search* (HSW) [17] measures, calculating the semantic distance in Wikipedia through a random walk guided by a web-based measure. This model is the first hybrid model to use web-based semantic proximity in Wikipedia and offers the benefit to provide several publications tracking the advancement and application of the model [18–23]. Moreover, the HSW model can be used and enhanced with any suitable proximity measure.

C. Flickr Distance

Flickr distance (FD) [24] is a model to measure the relationship between semantic concepts in a visual domain like Flickr.

The FD between two concepts x and y are measured by the average square root of the Jensen-Shannon divergence between the corresponding latent topic-based visual language model.

The major drawback of the Flickr distance is that it has not been validated with other datasets.

$$D_{flickr}(x,y) = \sqrt{\frac{\sum_{i=1}^{K}\sum_{j=1}^{K} D_{JS}(P_{z_i}x||P_{z_j}y)}{K^2}} \tag{3}$$

where

$$D_{JS}(P_{Z_j}x||P_{Z_j}x) = \frac{1}{2}D_{KL}((P_{Z_i}x||M) + \frac{1}{2}D_{KL}(P_{Z_j}x||M)). \tag{4}$$

The Kullback-Leibler divergence, expressed by:

$$D_{KL}(Q||R) = \sum\nolimits_i Q(i)\log\frac{Q(i)}{R(i)}, \tag{5}$$

is a standard measurement of the difference between two probability distributions Q and R. in the Flickr context, it can be seen as a measurement of the mutual entropy between two visual language models.

Moreover, computing the distance requires a computationally demanding process on the images content to calculate the probability distributions, which is hardly applicable to large and dynamical domains.

D. Structural Distances for Proximity Prediction

Link Prediction (LP) [25] is a method to establish the strength of ties in a graph: the more the strength, the more robust the ties among potential collaborators, e.g., co-authors in a bibliometric network [26]. The structural, i.e., the topological, similarity is a suitable candidate method for link prediction.

Common Neighbours-based (CN) rankings (e.g., Jaccard, Adamic-Adar [27]) represent a class of similarity measures for link prediction, which efficiently evaluate the likelihood of a new link based on the neighbors frontier of the already existing nodes. The CN proximity measures are more efficient than others (e.g., preferential attachment, path-based distances), but they have the weakness of returning a large 0-tail: [19, 20] a zero-rank value is given to all the links of pair of nodes, which have no common neighbors. Even if the rating is equal to 0, such links may be potentially suitable for link prediction.

Recent approaches for link prediction [28, 29] use topological (e.g., Adamic, Common Neighbours) and semantic similarity (e.g., Confidence, [30–32] PMI [33] and PMING [34–36]) together [37] or similarly, [9] to predict future links or infer existing unknown links from a social network or graph-based database [26]. Using the proper metrics, this approach allows shortening the 0-tail [10, 38–40].

Other approaches use a 2-steps or 3-steps traversal in order to navigate the network in short paths by different directions, in order to enhance the probability to find candidates for prediction.

3.2 Web-Based Semantic Similarity

The class of Web-Based Proximity Measures (WBPM) [3, 4, 41–43] is characterized by the probabilities of events used in their definitions, which can be approximated by frequencies drawn from popular web engines. Search engines are a accepted source of evidence about semantic affinity. The general idea is to use a search engine S as a black box to submit queries on a term or a set of terms.

In the following, we will use the term similarity and proximity as synonymous while, in other cases, when we will refer to distances. Simple arithmetic and normalization are used to transform a distance into proximity and vice-versa. We observe that WBPM can be equivalently based on probabilities or frequencies.

If the frequency $f(x) = |S(x)|$ denotes the cardinality of the results of a query to S based the search term x, and N is the total number of objects indexed by the S, $P(x) = f(x)/N$.

Different similarity measures have been experimented to assess the effectiveness of WBPM.

In the following, the WBPMs considered in our set-distance experiments are detailed.

A. Confidence and Mutual Confidence

Confidence (Conf) is a statistical measure applied in rule mining.

Given a rule $x \rightarrow y$, given the number of transactions that contain x, confidence calculates the percentage of transactions that also contain y:

$$Conf(x \rightarrow y) = \frac{P(x \wedge y)}{P(x)} = \frac{f(x \wedge y)}{f(x)} \tag{6}$$

Confidence cannot be used in a group distance, because it is not symmetric. Therefore, Mutual Confidence [22–24] is defined with the idea is to calculate the confidence of a term pair in both directions:

$$MC(x \leftrightarrow y) = \frac{1}{2}\frac{P(x)P(x \wedge y) + P(y)P(x \wedge y)}{P(x)P(y)} \tag{7}$$

B. Jaccard Similarity Measure

Jaccard similarity [19, 31, 32] is defined as a similarity ratio:

$$Jaccard = \frac{a}{a+b+c} \tag{8}$$

where a, b, and c represent the documents frequency of two terms w_1 and w_2 in a search:

$$a = w_1 \wedge w_2 \quad b = w_1 \wedge \neg w_2 \quad c = w_2 \wedge \neg w_1$$

C. Dice Similarity Measure

Dice similarity measure, [44] i.e. *Czekanowski* or *Sorenson*, gives more emphasis to the presence of common interactions between the compared sets, with respect to Jaccard:

$$Dice = \frac{2a}{2a + b + c} \tag{9}$$

where a, b, and c are defined as in the Jaccard measure.

3.3 Set Semantic Similarity

Our web-based set similarity [45–47] is defined on the elements extracted with a simple pre-processing from the context of a pair of images I_i and I_j. The set similarity scheme composes the elementary web-based similarities (or distances) d of the elements in the first and second sets of concepts.

$G_i = \{T_i1, T_i2, \ldots, T_im\}$ and $G_j = \{T_j1, T_j2, \ldots, T_jm\}$ is the group of concept elements (e.g. terms, hashtags, categories) associated to the images I_i and I_j.

The *Document Concept Set* (DCS) distance, measuring the mutual similarity of images I_i and I_j., is then defined by:

$$Dcs(I_i, I_j) = AVG_2\{AVG_1\{SEL \bigcup_{\substack{r\in 1,n \\ k\in 1,m}} d(T_{i_k}, T_{j_r})\}, AVG_1\{SEL \bigcup_{\substack{k\in 1,n \\ r\in 1,n}} d(T_{j_r}, T_{i_k})\}\} \tag{10}$$

where $SEL \in \{MAX, AVG, MIN\}$ and d is the web-based proximity or the composition of more web-based proximities. In our experiments, $d \in \{MC, Jaccard, Dice\}$.

As previously noticed, the DCS scheme defines a new class of semantic proximity measures, [26] where several elementary distances and a different composition operator generate a single semantic similarity function of the class.

4 Experiments

In our experiments, we demonstrated the applicability of the set distance for prosthetic knowledge on images in the Flickr repository, and then we designed a project to apply it to the exchange of learning objects between e-learning systems through an Artificial Intelligence-based assistant.

4.1 Image Set Similarity Experiments on Flickr Images

The Set Image Similarity model has been evaluated by experiments on a data set including 520 pairs of images from Flickr using our DCS [48].

The image tags annotated by Flickr users have been used as groups of concepts representing the knowledge associated with the image. Nine different instances of the DCS have been calculated varying SEL, with *SEL* ∈ *{MAX, AVG, MIN}* and *d* ∈ *{MC, Jaccard, Dice}*.

Table 1. A subset of the experiment results

image1	image2	HSE user score	D_{CS} MC	D_{CS} Jaccard	D_{CS} dice
1	2	0.4783	0.3980	0.5241	0.6309
3	4	0.5304	0.3372	0.4208	0.8384
5	4	0.3739	0.2956	0.1653	0.3431
6	4	0.4174	0.2885	0.8518	0.1335
2	7	0.3217	0.2779	0.5383	0.8821
8	9	0.2957	0.2593	0.4587	0.5877
1	10	0.4348	0.2493	0.7110	0.5320
11	6	0.7565	0.2374	0.2803	0.4891
12	13	0.3130	0.2288	0.4917	0.5219
4	14	0.1304	0.2266	0.6247	0.6996
4	15	0.1652	0.2105	0.3844	0.8873

A human similarity evaluation (HSE) of each pair of images has been gained as ground truth for evaluation, gathering the votes of 120 students of the Hong Kong Baptist University, who contributed to the tests on 100 casually mined pairs for each contributor. Each dyad of images has been evaluated giving a ranking score from *0* (=very different) to *5* (=very similar), based on subjective decision. The students' ranking scores have been averaged to acquire the HSE for each image dyad.

The DCS occurrences have been calculated using the statistics gave back from the search engines *Google* and *Bing*, and the search interface of *Wikipedia*. In order to assess the performance of the rank of image dyads formed by the different occurrences of DCS and HSE, we have calculated the Pearson's correlation between them. Table 1 shows a pre-normalization HSE, i.e., the contributor's average score for some image dyad of the tests before normalization; MC ranks the results.

The systematic evaluation of the experimental results shows that the composition scheme where SEL = MAX and d = Dice is the most effective among the class, compared to human evaluation. The best values of Pearson's correlation score for MAX-Dice is 0.732, while MAX-MC is 0.726. The other grouping of measures attained lower scores on this specific task, without noteworthy variances regarding the different search engines.

4.2 Searching e-Learning Objects with an AI-Based Assistant

Recently, Artificial Intelligence (AI) has been introduced to e-learning systems [5]. In our environment, the exchange of Linkable Learning Objects (LLO) [49, 50] can be used for a dialogue between Learning Systems to obtain information, especially with the use of semantic or structural similarity measures to enhance the classification.

Learning open-source platforms, e.g., GLOREP and Moodle, are particularly suitable to be enhanced with AI-based tools, e.g., an intelligent assistant to support information exchange. A feasible solution is to develop an assistant based on existing devices, e.g. Google Home, Amazon Alexia, Microsoft Cortana, which offer a set of API for voice control. The module *moodledata* can be used to manage the information exchange between learning systems, both in GLOREP and in Moodle.

In such a web-based environment, web-based similarity measures, which rely on search engines (e.g., Google, Bing) to obtain frequency values used for semantic proximity calculation, can be easily integrated into an AI to obtain a set of LLO to exchange among learning systems.

Figure 2 depicts the scheme of our AI-based LLO search system, showing the exchange of LLOs between Moodle and a human. The first set of terms is the expression pronounced by the human via the voice control system of the AI-assistant, and the second set of terms is each set of metadata included in the learning system database.

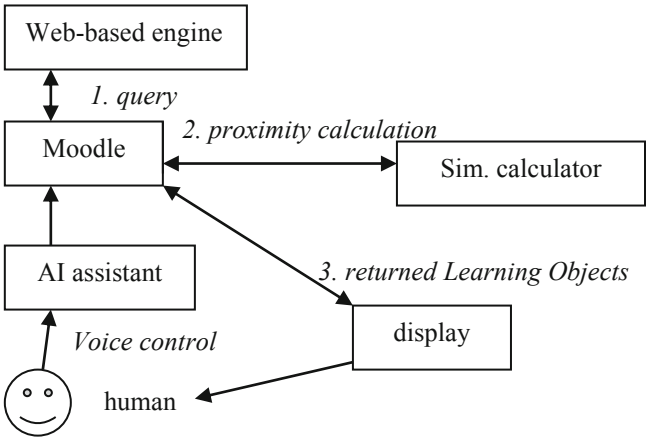

Fig. 2. AI-based LLO exchange

5 Conclusions and Future Work

Concept information can be expressed by text, images or general objects which semantic meaning is clear to a human in a specific cultural context. For a computer, when available, text with its semantics (e.g., metadata, comments, captions) can convey more precise meaning than images or general objects with low-level features (e.g.,

color distribution, shapes, sound peaks) to extract the concept underlying the object. We have shown how web-based proximity measures are sufficient for concept evaluation, exploiting statistical data provided by search engines on terms and expressions provided in texts associated with the object. We experimented the set image on the web-based distances Mutual Confidence, Dice and Jaccard with a selection scheme using Average, Maximum and Minimum on the Wikipedia search, Google and Bing. Experiments indicate that the best composition scheme is the one using Dice and MAX, compared to human evaluation on both the search engines compared to the Wikipedia search.

Furthermore, where Artificial Intelligence can be a support for impaired individuals, e.g., having disabilities related to vision and hearing, understanding the concept underlying an object can be critical for an intelligent artificial assistant. Also, for people neither mentally nor physically impaired, knowledge can be prosthetic in the sense of augmenting the knowledge through a device.

In this work we proposed to use the set semantic distance, already used in literature for semantic similarity measurement of web objects, as a tool for artificial assistants to support knowledge extraction; in other words, as prosthetic knowledge. As an example application, we designed a learning object selector, capable of selecting in a learning system the object that best suits a search made via an AI assistant.

Acknowledgments. The authors thank the students involved in the experiments, and the authors of previous works of the image set similarity project, cited in this paper, in particular Alfredo Milani, Clement H.C. Leung, Sheung Wai Chan, Marco Mencacci, Paolo Mengoni, and Simonetta Pallottelli.

References

1. Franzoni, V., Milani, A., Nardi, D., Vallverdú, J.: Emotional machines: the next revolution. Web Intell. **17**, 1–7 (2019)
2. Franzoni, V., Milani, A., Vallverdú, J.: Emotional affordances in human-machine interactive planning and negotiation. In: Proceedings of 2017 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2017 (2017)
3. Milani, A., Rajdeep, N., Mangal, N., Mudgal, R.K., Franzoni, V.: Sentiment extraction and classification for the analysis of users' interest in tweets. Int. J. Web Inf. Syst. **14**, 29–40 (2018)
4. Mudgal, R.K., Niyogi, R., Milani, A., Franzoni, V.: Analysis of tweets to find the basis of popularity based on events semantic similarity. Int. J. Web Inf. Syst. **14**, 438–452 (2018)
5. Franzoni, V., Mengoni, P., Milani, A.: Dimensional morphing interface for dynamic learning evaluation. In: Information Visualisation - Biomedical Visualization, Visualisation on Built and Rural Environments and Geometric Modelling and Imaging, IV 2018 (2018)
6. Gervasi, O., Franzoni, V., Riganelli, M., Tasso, S.: Automating facial emotion recognition. Web Intell. **17**, 17–27 (2019)
7. Zhang, L., Ma, W.-Y., Li, X., Lin, F., Chen, L.: Image annotation by large-scale content-based image retrieval (2007)
8. Budanitsky, A., Hirst, G.: Evaluating WordNet-based measures of lexical semantic relatedness. Comput. Linguist. **32**, 13–47 (2006)

9. Milani, A., Franzoni, V., Biondi, G., Li, Y.: Integrating binary similarity measures in the link prediction task (2019)
10. Franzoni, V., Chiancone, A., Milani, A.: A multistrain bacterial diffusion model for link prediction. Int. J. Pattern Recognit Artif Intell. **31**, 1759024 (2017)
11. Budanitsky, A., Hirst, G.: Semantic distance in WordNet : an experimental, application-oriented evaluation of five measures. In: Workshop on WordNet and Other Lexical Resources, vol. 2 (2001)
12. Pedersen, T., Patwardhan, S., Michelizzi, J.: WordNet:: similarity: measuring the relatedness of concepts. In: Demonstration Papers at HLT-NAACL 2004 (2004)
13. Lin, J.: Divergence measures based on the shannon entropy. IEEE Trans. Inf. Theory **37**, 145–151 (1991)
14. Strube, M., Ponzetto, S.P.: WikiRelate! Computing semantic relatedness using Wikipedia. Am. Assoc. Artif. Intell. **6**, 1419–1424 (2006)
15. Gabrilovich, E., Markovitch, S.: Computing semantic relatedness using wikipedia-based explicit semantic analysis. In: IJCAI International Joint Conference on Artificial Intelligence (2007)
16. Milne, D.: Computing semantic relatedness using Wikipedia link structure. In: Work (2007)
17. Franzoni, V., Milani, A.: Heuristic semantic walk: browsing a collaborative network with a search engine-based heuristic (2013)
18. Franzoni, V., Milani, A.: A pheromone-like model for semantic context extraction from collaborative networks. In: Proceedings of 2015 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, WI-IAT 2015 (2016)
19. Franzoni, V., Mencacci, M., Mengoni, P., Milani, A.: Semantic heuristic search in collaborative networks: measures and contexts. In: Proceedings of 2014 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology - Workshops, WI-IAT 2014 (2014)
20. Franzoni, V., Milani, A.: Heuristic semantic walk for concept chaining in collaborative networks. Int. J. Web Inf. Syst. **10**, 85–103 (2014)
21. Franzoni, V., Mencacci, M., Mengoni, P., Milani, A.: Heuristics for semantic path search in Wikipedia (2014)
22. Franzoni, V., Milani, A.: Semantic context extraction from collaborative networks. In: Proceedings of the 2015 IEEE 19th International Conference on Computer Supported Cooperative Work in Design, CSCWD 2015 (2015)
23. Pallottelli, S., Franzoni, V., Milani, A.: Multi-path traces in semantic graphs for latent knowledge elicitation. In: Proceedings of the International Conference on Natural Computation (2016)
24. Wu, L., Hua, X.-S., Yu, N., Ma, W.-Y., Li, S.: Flickr distance: a relationship measure for visual concepts. IEEE Trans. Pattern Anal. Mach. Intell. **34**, 863–875 (2012)
25. Liben-Nowell, D., Kleinberg, J.M.: The link-prediction problem for social networks. JASIST **58**, 1019–1031 (2007)
26. Franzoni, V., Lepri, M., Li, Y., Milani, A.: Efficient graph-based author disambiguation by topological similarity in DBLP. In: Proceedings of 2018 1st IEEE International Conference on Artificial Intelligence and Knowledge Engineering, AIKE 2018 (2018)
27. Adamic, L.A., Lento, T.M., Adar, E., Ng, P.C.: Information evolution in social networks. In: WSDM, pp. 473–482. ACM (2016)
28. Hoffman, M., Steinley, D., Brusco, M.J.: A note on using the adjusted Rand index for link prediction in networks. Soc. Netw. **42**, 72–79 (2015)
29. Han, S., Xu, Y.: Link prediction in microblog network using supervised learning with multiple features. JCP **11**, 72–82 (2016)

30. Biondi, G., Franzoni, V., Li, Y., Milani, A.: SEMO: a semantic model for emotion recognition in web objects. In: ICCSA. Springer, Heidelberg (2017)
31. Franzoni, V., Biondi, G., Milani, A.: A web-based system for emotion vector extraction. In: Gervasi, O., et al. (eds.) ICCSA 2017. LNCS, vol. 10406, pp. 653–668. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-62398-6_46
32. Huang, A.H., Yen, D.C., Zhang, X.: Exploring the potential effects of emoticons. Inf. Manag. **45**, 466–473 (2008)
33. Turney, P.D.: Mining the web for synonyms: {PMI-IR} versus {LSA} on {TOEFL}. CoRR. cs.LG/0212 (2002)
34. Franzoni, V., Milani, A.: PMING distance: a collaborative semantic proximity measure. In: Proceedings of 2012 IEEE/WIC/ACM International Conference on Intelligent Agent Technology, IAT 2012 (2012)
35. Biondi, G., Franzoni, V., Li, Y., Milani, A.: Web-based similarity for emotion recognition in web objects. In: Proceedings of 9th IEEE/ACM International Conference on Utility and Cloud Computing, UCC 2016 (2016)
36. Franzoni, V., Milani, A., Biondi, G.: SEMO: a semantic model for emotion recognition in web objects. In: Proceedings of 2017 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2017 (2017)
37. Franzoni, V., Milani, A.: Structural and semantic proximity in information networks. In: Gervasi, O., et al. (eds.) ICCSA 2017. LNCS, vol. 10404, pp. 651–666. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-62392-4_47
38. Chiancone, A., Franzoni, V., Li, Y., Markov, K., Milani, A.: Leveraging zero tail in neighbourhood for link prediction. In: Proceedings of 2015 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, WI-IAT 2015 (2016)
39. Chiancone, A., Milani, A., Poggioni, V., Pallottelli, S., Madotto, A., Franzoni, V.: A multistrain bacterial model for link prediction. In: Proceedings of International Conference on Natural Computation (2016)
40. Chiancone, A., Franzoni, V., Niyogi, R., Milani, A.: Improving link ranking quality by quasi-common neighbourhood. In: Proceedings of 15th International Conference on Computational Science and Its Applications, ICCSA 2015 (2015)
41. Franzoni, V., Milani, A.: A semantic comparison of clustering algorithms for the evaluation of web-based similarity measures. In: Gervasi, O., et al. (eds.) ICCSA 2016. LNCS, vol. 9790, pp. 438–452. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-42092-9_34
42. Franzoni, V., Li, Y., Mengoni, P.: A path-based model for emotion abstraction on facebook using sentiment analysis and taxonomy knowledge. In: Proceedings of 2017 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2017 (2017)
43. Leung, C.H.C., Li, Y., Milani, A., Franzoni, V.: Collective evolutionary concept distance based query expansion for effective web document retrieval. In: Murgante, B., et al. (eds.) ICCSA 2013. LNCS, vol. 7974, pp. 657–672. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39649-6_47
44. Brown, A., Randall, S., Ferrante, A., Boyd, J.: Partial Agreements in Probabilistic Linkages. Int. J. Popul. Data Sci. **3**, 293 (2018)
45. Franzoni, V., Leung, C.H.C., Li, Y., Mengoni, P., Milani, A.: Set similarity measures for images based on collective knowledge. In: Gervasi, O., et al. (eds.) ICCSA 2015. LNCS, vol. 9155, pp. 408–417. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-21404-7_30
46. Zhang, J., Zhou, Q., Zhuo, L., Geng, W., Wang, S.: A CBIR system for hyperspectral remote sensing images using endmember extraction. IJWPRAI **31**(4) (2016)

47. Franzoni, V., Milani, A., Pallottelli, S., Leung, C.H.C., Li, Y.: Context-based image semantic similarity. In: 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2015 (2016)
48. Chan, S.W., Franzoni, V., Mengoni, P., Milani, A.: Context-based image semantic similarity for prosthetic knowledge. In: Proceedings of 2018 1st IEEE International Conference on Artificial Intelligence and Knowledge Engineering, AIKE 2018 (2018)
49. Franzoni, V., Tasso, S., Pallottelli, S.: Sharing linkable learning objects between a content management system and a learning management system with the use of metadata and a taxonomy assistant for categorization. In: LNCS, ICCSA 2019 (2019)
50. Tasso, S., Pallottelli, S., Gervasi, O., Sabbatini, F., Franzoni, V.: Cloud and local servers into a federation of learning object repositories. In: ICCSA 2019, LNCS. Springer, Heidelberg (2019)