

Sujet de thèse •  
 PhD Subject Title:

## Analogical proportions within web semantic languages

 Résumé •  
 Abstract:

### Context:

An analogical proportion binds together **four** objects A, B, C, D of the same type in a proposition of the form: “A **is to** B **as** C **is to** D”. They allow for expressing the identity of the ratio between two pairs of elements. Typical illustrations of this notion could be, in natural language: “Calf is to cow as foal is to mare” [6], “Aurochs is to ox as mammoth is to elephant”, “electrons are to nucleus as planets are to sun” [15]. This type of relationship expresses that what distinguishes A and B (and what they share) is comparable to what distinguishes C and D (and what they share). The above examples show the diversity (and potential complexity) of the possible semantics of the “is to” connective involved in an analogical proportion.) When the objects A and B (resp. C and D), represent the same entities at different moments/states of their life (for example, A and B describe the same place at two distinct periods), the analogical proportion may express similar variations of two distinct situations. Then, analogical proportions can parallel historical or socio-economical facts and environmental situations. To the best of our knowledge, analogical proportions have been mainly investigated in artificial intelligence, and notably for classification (as for example in [2]). The aim of this project is to investigate the notion of *analogical proportion* in the context of the semantic web, in order to propose new tools for reasoning and data querying.

### Challenges:

All the above examples highlight the necessity to have a **clear semantic description** of the environmental objects for discovering rich parallels between objects. Consequently, we will **instantiate the study of analogical proportions** in languages of the semantic web promoted by the W3C, and notably, **the DL-Lite family languages** (Calvanese, JAR 2007). These languages, based on description logics (Baader, Handbook), have the advantages to be able to represent sample data as list of property-values pairs as well as semantically annotated data. At last, they will supply a first simple framework to define tools for discovering analogical proportions occurring in semantic data set. These results can be further extended to more expressive languages.

The second objective is the definition of **new types of queries**. Then, we aim at devising four kinds of querying. The **blind research** is the research of quadruplets that are bound by a analogical relationship, which implements the ideas developed above, in the sense that the query mines the database or the knowledge base to find such types of objects. In order to reduce the research space and to provide a simplified querying interface to users, the **directed querying** tool will allow the user to research for objects bound by a particular type of analogical proportion, selected from a suggested list (e.g., filiation, particular schema of evolution). As previously mentioned, the analogical proportion relationship may denote a filiation relationship, or the sharing of properties, or the adding and the deletion of some properties. More concretely, one can be interested in retrieving areas or countries having undergone the same changes. This querying tool requires to semantically **characterize a typology of the analogical relationships**, and when possible, deduce their own axiomatic. The types of the analogical proportions will provide a basis for querying data and for the definition of index to improve the execution times of the blind querying. The **querying by example** tool must search, given two objects, for 2-uplets that are logically proportional with the two objects. Indeed, it could be very convenient for users to query the system by giving the concrete example of a new situation whose requires a better understanding. The user gives as input the situation to observe, in order to mine situations with which they are bound by an analogical proportion relationship. Another interesting tool is that of **checking the veracity of an assertion and explaining the result**. For example, one can imagine that user has an idea of a situation that he suspects to follow another well-known situation.

He asks the system for knowing if the parallel is relevant or not and why.

### Positionning

Analogical proportions have been also widely used in artificial intelligence, e.g., for linguistics analysis tasks [12], [13], [15], [14]. More recently, it has been proven that analogical propositions offer an interesting framework for classification ([10], [17]) and for solving IQ tests [7]. Moreover, a propositional logic modelling of analogical proportions has been proposed and then extended to fuzzy logics [9].

A more restrictive set of recent works is more related to our concern. Two works [2] and [18] are interested in the establishment of links between the setting of formal concept analysis and that of analogical proportions. In such a setting, it is possible to discover analogical proportions that hold in the underlying lattice. A last work [16], the only one to our knowledge that explicitly shares our objective of discovering analogical proportions, aims at enumerating all analogical proportions occurring between Chinese characters. Each character is modelled as a vector of features and a feature is encoded by an integer. This work, however, must be extended in order to comply with our framework based on DL-Lite family of languages. Indeed, these latter works use frameworks much less expressive than those of ontological languages such as DL-Lite family of description logics, at the heart of our study. Indeed, [2] exploits the formal concept analysis (FCA) setting that focuses only on the representation of Boolean properties and [16] uses vectors of integers.

Finally, let us mention a recent work concerning analogical reasoning in description logics [1]. This work focuses on the induction of new instances and on the classification of instances, by following a KNN approach and by using a dissimilarity measure defined in terms of the semantic descriptions of concepts and instances, and assuming the existence of a probability distribution for the concepts of the knowledge base. The authors are not interested in the problem of knowledge discovery, and of the verification of assertions.

To the best of our knowledge, the notion of analogical proportion has not plainly been exploited in the setting of knowledge discovery. It is worth noticing that our concern in this project is to model the analogy in the formal framework of the DL-Lite family of languages. The proposed definition of analogical proportions will use, as a semantic context, a knowledge base. Another objective is to propose new services as discovery of analogical proportions, verification of analogical assertions and explanation. Then, among all the cited works, none tackles the issue of modelling analogy in the DL-Lite family of languages, nor the definition of a query language based on analogy for knowledge discovery in the semantic web context. It would be of interest to relate the obtained modelling with propositions made in cognitive sciences that use other structural frameworks.

A perfect applicant should have strong background in Computer Sciences, and should be inquisitive, autonomous, dynamic and interested both in theoretical and practical aspects of the subject. The student will be part of the Shaman team of the IRISA, located at Lannion.

**Dpt scientifique •**  
Scientific department:

D7- Gestion des données et de la connaissance  
Data and Knowledge Management

**Equipe projet •**  
Research team:

Shaman (DKM Department of IRISA)  
<http://www-shaman.irisia.fr/>

**Directeur de thèse •**  
 PhD Director:

François GOASDOUE, PR

**Encadrant(s) •**  
 PhD supervisors:

 François GOASDOUE (PR IRISA/DKM/SHAMAN)  
 Hélène JAUDOIN (MCF IRISA/DKM/SHAMAN )

**Contact(s) :**

To apply for this position please send your Cover Letter, CV as well as recommendation letters to:

François Goasdoué &lt;fg@irisia.fr&gt; and Hélène Jaudoin&lt;hjaudoin@irisia.fr&gt;

**Début des travaux •**  
 Work start date:

October 2014

**Lieu •**  
 Place

 IRISA – Lannion ([http://www.ville-lannion.fr/en/accueil\\_en.html](http://www.ville-lannion.fr/en/accueil_en.html)), Pink Granite Coast of Brittany, France.

**Bibliographie •**  
 References:

- [1] d'Amato, C., Fanizzi, N., Esposito, F. : Analogical reasoning in description logics. In Paulo Cesar G. Costa, Claudia d'Amato, Nicola Fanizzi, Kathryn B. Laskey, Kenneth J. Laskey, Thomas Lukasiewicz, Matthias Nickles, and Michael Pool, editors, *Uncertainty Reasoning for the Semantic Web I*, volume 5327 of *Lecture Notes in Computer Science*, pages 330–347. Springer Berlin Heidelberg, 2008.
- [2] Barbot, N., Miclet, L., Prade, H.: Proportions analogiques et factorisation de l'information dans les treillis distributifs. In: *Journées d'Intelligence Artificielle Fondamentale. GDR I3*, Aix en Provence, France, 2013.
- [3] Meriam Bayouhd, Henri Prade, and Gilles Richard. Evaluation of analogical proportions through kolmogorov complexity. *Knowl.-Based Syst.*, 29 :20–30, 2012.
- [4] Sabri Bayouhd, Laurent Miclet, and Arnaud Delhay. Learning by analogy : A classification rule for binary and nominal data. In Manuela M. Veloso, editor, *IJCAI*, pages 678–683, 2007.
- [5] Anouar Ben Hassena and Laurent Miclet. Analogical learning using dissimilarity between tree-structures. In *European Conference on Artificial Intelligence*, pages 1039–1040, Lisbon, Portugal, 2010.
- [6] Antoine Cornuejols and Laurent Miclet. *Apprentissage artificiel : Concepts et algorithmes*, 2ème édition. Eyrolles, 2010.
- [7] Correa, W., Prade, H., Richard, G.: When intelligence is just a matter of copying. In: *Eur. Conf. on Artificial Intelligence*, pp. 276–281, IOS Press (2012)
- [8] Dedre Gentner, Keith J. Holyoak, and Boicho N. Konikov. *The analogical mind : Perspectives from cognitive science*. MIT Press, 2001.
- [9] Prade, H. et G. Richard. Analogical proportions and multiple-valued logics. In *Proc. of the 12th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'13)*, LNCS, vol. 7958 , pp. 497–509.
- [10] Henri Prade, Gilles Richard, Bing Yao. Enforcing regularity by means of analogy-related proportions -- A new approach to classification. Dans : *International Journal of Computer Information Systems and Industrial Management Applications*, Dynamic Publishers, Atlanta - USA, Vol. 4, p. 648-658, 2012.
- [11] Henri Prade and Gilles Richard. From analogical proportion to logical proportions. *Logica Universalis*, 7(4) :441–505, 2013.
- [12] Stroppa, N., Yvon, F.: An analogical learner for morphological analysis. *Proc. Conf. Comput. Natural Language Learning*, pp. 120–127. (2005)

- [13] Stroppa, N., Yvon, F.: Du quatrième de proportion comme principe inductif : une proposition et son application à l'apprentissage de la morphologie. *Traitement Automatique des Langues*, 47(2), 1–27 (2006)
- [14] Lepage, Y. De l'analogie rendant compte de la commutation en linguistique. HDR. University of Joseph Fourier (Grenoble, France), 2003.
- [15] Lepage, Y.: Analogy and formal languages. *Elec. Notes Theo. Comp. Sci.*, 53 (2001)
- [16] Lepage, Y.: (Re-)discovering the graphical structure of Chinese characters. Workshop SAMAI 2012, co-located with ECAI2012, Montpellier, France, 2012
- [17] Miclet, L., Bayouhd, S., Delhay, A.: Analogical dissimilarity: Definition, algorithms and two experiments in machine learning. *J. Artif. Intell. Res. (JAIR)* 32, 793-824, 2008
- [18] Miclet, L., Prade, H., Guennec, D. : Looking for Analogical Proportions in a Formal Concept Setting In: *Proceedings of the 8th International Conference on Concept Lattices and Their Application*, 17-20 October, Nancy, France, pages 295–307, 2011.

Mots clés •  
Keywords

Analogy, Semantic Web languages, Artificial Intelligence