

Emerging Discursive Formations, Folksonomy and Social Semantic Information Spaces (SSIS): The Contributions of the Theory of Integrative Levels in the Studies carried out by the Classification Research Group (CRG)

Maria Aparecida Moura

School of Information Science, Federal University of Minas Gerais, Brazil,
Av. Antônio Carlos, 6627, Pampulha 31270-901, Belo Horizonte, MG – Brazil
<mamoura@eci.ufmg.br>, <cidamoura@gmail.com>



Maria Aparecida Moura holds a degree in library science, has a Ph.D. in communication and semiotics and a post doctorate in cognitive semiotics and new media. She is a titular professor from the School of Information Studies of the Federal University of Minas Gerais (ECI/UFMG), and is the Director of Information Governance in UFMG. Her range of interests includes semiotics and collaborative media, social and theoretical foundations of information and knowledge organization in collaborative environments.

Moura, Maria Aparecida. **Emerging Discursive Formations, Folksonomy and Social Semantic Information Spaces (SSIS): The Contributions of the Theory of Integrative Levels in the Studies carried out by the Classification Research Group (CRG).** *Knowledge Organization*. 41(4), 304-310. 9 references.

Abstract: This paper focuses on the discursive formations emerging from the Social Semantic Information Spaces (SSIS) in light of the concept of emergence in the theory of integrative levels. The study aims to identify the opportunities and challenges of incorporating epistemological considerations in the act of acquiring knowledge into the consolidation of knowledge organization and mediation processes and devices in the emergence of phenomena. The goal was to analyze the effects of that concept on the actions of a sample of researchers registered in an emerging research domain in SSIS in order to understand this type of indexing done by the users and communities as a classification of integrating levels. The methodology was established by triangulation through social network analysis, consensus analysis and archaeology of knowledge. It was possible to conclude that there is a collective effort to settle a semantic interoperability model for the labeling of contents based on best practices regarding the description of the objects shared in SSIS.

Received: 12 June 2014; Accepted: 14 June 2014

Keywords: discursive formations, folksonomy, Social Semantic Information Spaces, integrative levels, Classification Research Group

1.0 Introduction

We must question those ready-made syntheses, those groupings that we normally accept before any examination, those links whose validity is recognized from the outset. We must oust those forms and obscure forces by which we usually link the discourse of one man with that of another; they must be driven out from the darkness in which they reign—Foucault

The expansion of the Internet, widening and diversifying the social representations offered by web users, has been requiring more expedited solutions, able to incorporate the emerging discursive formation by means of different devices available on the Internet and dedicated to sharing information. Although the theoretical studies and practical experiences used to focus on the material nature of information circulation devices, nowadays, given the exchange between the social roles of actors and information receivers, it is possible to state that the focus has mi-

grated to sharing, production, management and mediation. This trend is pursuant to the fact that the informational subject has become a source of information in itself, as the actor is required to comprehend, systematize, manipulate and dynamize informational flows in Social Semantic Information Spaces (SSIS).

The SSIS are characterized as technological environments that link people, knowledge and computer resources. Because they are grounded on social connectivity and semantic technologies, they stimulate the coexistence of specialized communities, organized around information sharing by means of themes and concepts. These semantic spaces arise in the form of private or public social platforms, inside which profiles of individual researchers and investigation groups are generated, sharing different information quality standards based on distinct research interests and insights about the world.

Within this context, the term folksonomy was created by Vander Wal (2004), as a neologism associating the term "folk" with the term "taxonomy," with the purpose of emphasizing classification agreements on the content shared in digital environments. The purpose of this tool is to generate taxonomies used to categorize and retrieve the contents resulting from users' experiences and social networks in the web. The term folksonomy carries additional concepts, such as: ethnic classification, collaborative labeling, social bookmarking, social labeling, narrow folksonomy, popular taxonomy, and social classification, among others.

Here we assess the theoretical aspects of the concept of "emergence" within the theory of integrative levels and their development in the studies of the Classification Research Group (CRG). The goal was to analyze the effects of that concept on the actions of a sample of researchers registered in an emerging research domain in the website Academia.edu (<http://academia.edu/>), in order to understand this type of indexing done by users and communities as a classification of integrating levels. Our main challenge was to understand how researchers' representations shared in SSIS affected the linguistic behavior of the users in contexts of scientific information sharing. How, and to what extent, do the labels shared in such contexts match those in the systems of information representation adopted in instances of access mediation?

2.0 Emerging discursive formations

According to Foucault (2010), discourse is a set of statements that may belong to different fields but are nevertheless submitted to common functional rules (clinical, medical, psychiatric speech). Discourses should be handled in the instance in which they are played, once they do not represent a place of ease where other questions may arise

(structure, occurrence, systematics and transformations). For Foucault, discursive formation (2010, 133) is:

A body of anonymous, historical rules, always determined in time and space that have defined for a given period, and for a given social, economic, geographical or linguistic area, the conditions of operation of the enunciative function.

Contemporary discursive formations had inevitably been consolidated under the aegis of information and knowledge disclosure and sharing. The information system of the objects is permanently modified and slowly deposited into (medical, economic, social) discourse. Pursuant to that, the conceptual systems produced in such contexts are oftentimes heterogeneous or incompatible derivations, making it always necessary to analyze the way concepts arise and their disclosure in a given context. Understanding discursive formation, within the context, may help to identify the structural relationships and, consequently, support the acquisition of strong clues about the proposition of relationship systems that are better associated with emerging knowledge.

For the knowledge organization area, proposing information retrieval instruments in context without understanding the dynamics of discursive formation in a given domain of knowledge has become even more complex. This happens mainly because the concepts refer to the description of a field where the principles are set out. These are formed as a bundle of relations (not an isolated object, an individual work or an area of knowledge at a given moment), where the context, the discursive regularities and coercion, the theoretical choices and the historicity are taken into consideration, and happenings, transformations, mutations and processes are articulated.

Discursive formation is voiced at three levels: surfaces of emergence, authorities of delimitation and grids of specification. Surfaces of emergence highlight the circumstances of instantiation, individual differences and degrees of rationale, conceptual codes and the types of theories. The authorities of delimitation refer to the identification of institutions that adjust the domain and are responsible for distinguishing, assigning, naming and implementing the signification processes as object. Finally, the grids of specification refer to the grouping schemes and classifications adopted in a specific context of discursive formation.

Foucault (2010, 28-29), however, affirms that, in the scientific context:

There exists another principle of rarefaction of a discourse, complementary to the first, to a certain extent: the author. Not, of course, the author in the sense of the speaking individual who pronounced or

wrote a text, but in the sense of a principle of grouping of discourses, conceived as the unity and origin of their meanings, as the focus of their coherence. This principle is not everywhere at work, nor in a constant manner: there exists all around us plenty of discourses which circulate without deriving their meaning or their efficacy from an author to whom they could be attributed: everyday remarks, which are effaced immediately; decrees or contracts which require signatories but no author; technical instructions which are transmitted anonymously. But in the domains where it is the rule to attribute things to an author – literature, philosophy, science—it is quite evident that this attribution does not always play the same role. In the order of scientific discourse, it was indispensable, during the Middle Ages, that a text should be attributed to an author, since this was an index of truthfulness. A proposition was considered as drawing even its scientific value from its author.

From a philosophical point of view, emergence is the process by which a phenomenon arises from a combination of causes, while it may not be considered a result of its respective effects.

For Goldstein (1999, 50), emergence refers to the instantiation of innovative and coherent structures, standards and properties during the process of self-organization in complex systems. Its major characteristics are:

- Coherence or co-relation: emergents appear as integrated wholes that tend to maintain some sense of identity over time. That coherence spans and correlates the separate lower level components into a higher level unit.
- Global or macro level: since coherence represents a correlation that spans separate components, the locus of emergent phenomena occurs at global or macro level;
- Dynamic level: emergent phenomena are not pre-given wholes but arise as a complex system that evolves over time. As a dynamical construct, emergence is associated with the arising of new attractors in dynamical systems (i.e., bifurcation).
- Ostensive level: emergents are recognized by showing themselves, i.e., they are ostensively recognized.

Emergence is an important concept in the theory of integrative levels, which influenced the studies of the CRG.

3.0 The Classification Research Group and the theory of integrative levels

According to Gnoli (2011, 30), the structures of knowledge organization involve influences deriving from different factors. Among these factors, Gnoli highlights the ontological, epistemological and pragmatic dimensions. The ontological dimension deals with actual phenomena that are studied. The epistemological dimension deals with the purposes, the cultural, disciplinary and methodological aspects that are considered and discussed in the organization of knowledge. Finally, the pragmatic dimension involves material aspects related to documents, collections and information users.

Founded in 1948-1952, the CRG dedicated itself to the study of classification systems aiming at contributing to the development of documental research. The focus of its studies involved, above all, the study of a multiple sided assessment of the relational operators and the theory of integrative levels. According to Spiteri (1998), the theoretical work of the CRG embraced the study of facet analysis, relational operators and the theory of integrative levels. Employing the theory of integrative levels, the CRG construed the process of indexing as a complex and ontological relationship in the information producing environment and in the context of research. That importance given to the context of emergence of phenomena contrasted with the centrality of literary warrant adopted at that time.

Given the wide experience of the members of the CRG in the treatment of specialized information in different contexts, the group disagreed with the assumption of a finite universe of knowledge underlying the most traditional classification systems adopted at that time. The major limits of the classification systems referred to the difficulty to maintain and update systems due to changes in the production of knowledge, notably related to interdisciplinary and integrating dialogues pursuant to these processes.

As early as 1950, Farradane postulated on the support of epistemological considerations about the very process of knowledge acquisition (Rivier 1992, 63). From that point of view, the group proposed the establishment of semantic relations *a posteriori*, and received harsh criticism for abandoning, to a certain extent, Ranganathanian principles.

According to Foskett (1962), *apud* Spiteri (1995), the appropriation of the theory of integrative levels carried out by the CRG presented the following interpretation:

The world of things develops from the simple towards the complex by accumulation of new and divergent properties and that at certain points changes occur which transform the “entity” from a member of one group or class into a member of a

new group. The new entity has properties of its own, characteristic of the new level of organization within it, and it behaves in a similarly new and characteristic manner.

In accordance with Spiteri (1995), influenced by James Feiblemann (1965), the CRG proposed a set of categories that could apply to the entities, namely: physical entities, chemical entities, non-living heterogeneous entities, artifacts, biological entities, men and mentefacts. Currently, the possibilities to identify, monitor, record and segment the informational practices and the distinct user profiles in digital collaborative environments encourages the recovery of issues proposed by the CRG, with the purpose of identifying the objective potentials of incorporating the epistemological considerations of the process of knowing, organizing and mediating information upon the emergence of the phenomena.

4.0 The research

The systematic approach of human-social phenomena involves, in accordance with Domingues (2004), description, explanation and interpretation based on abstractions of the real world. Pursuant to that, the qualitative research developed had the purpose of describing, explaining and interpreting a contemporary phenomenon within the context of knowledge organization. In order to achieve that goal, the group chose to triangulate the methods and theories according to which the theory of

integrative levels was articulated, as well as the archaeology of knowledge, with the purpose of understanding the process of signification in digital environments and its repercussion in the representation and retrieval of information in dynamic contexts. By means of netography and the analysis of social networks (ARS), academic profiles, social aggregations and classifying practices of the Academia.edu site, users were identified and analyzed based on the emergent themes e-Science, e-Research, cyber-infrastructure, and science 2.0.

From a theoretical point of view, e-Science is the term that represents the new trend of scientific issues, according to Hey and Trefethen (2008, 15):

The next generation of e-Science problems range from the simulation of complex engineering and biological systems to research in bio-information technology, protoemics and pharmacogenetics. In many of these instances researchers need to combine the expertise of other research groups and access specialized resources often distributed across the globe Similarly in the social sciences the data sets that need to be analyzed are frequently so large and widely distributed that both the memory storage as well as the computational power of individual workstations are inadequate and the use of distributed computing resources will become the norm.

The website Academia.edu (Figure 1) is an open science digital platform dedicated to sharing articles and docu-

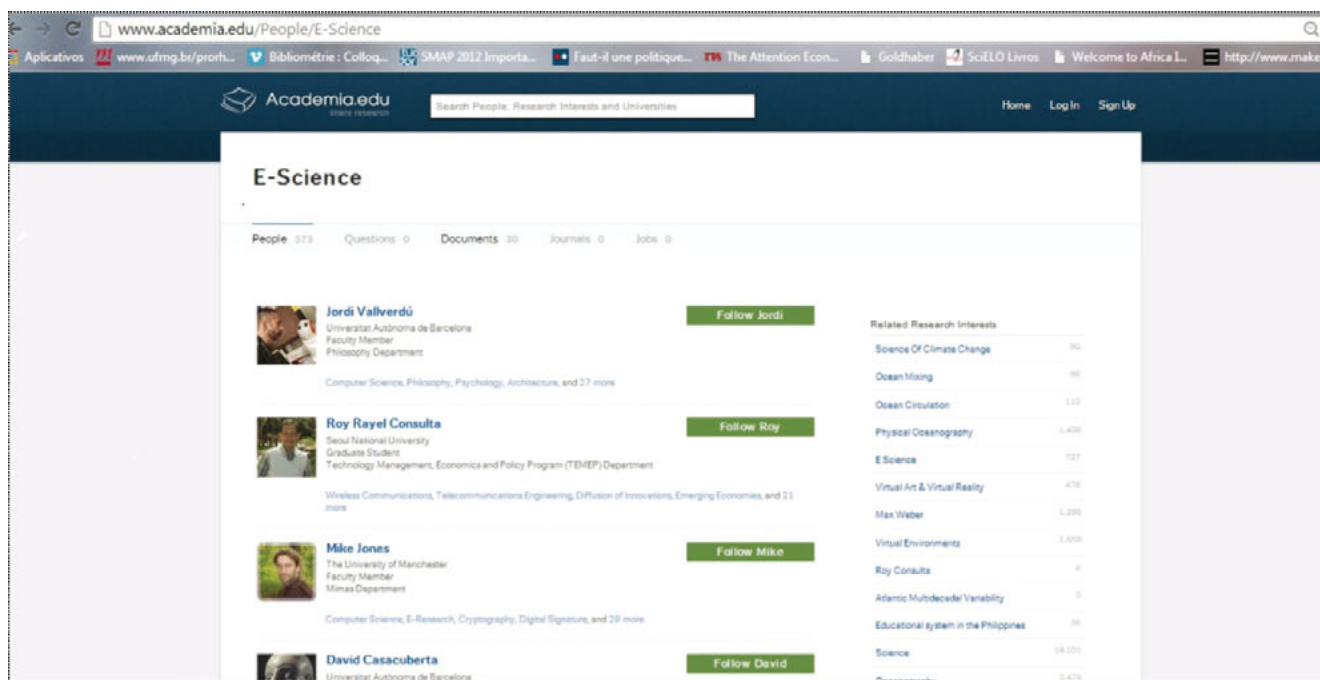


Figure 1. Academia.edu

ments developed at different stages of the research among academic practitioners. This platform was launched in September, 2008. With that objective, the researchers are able to create academic profiles in the social platform and offer institutional information focused in their research interests. That information generates an individual site with the researcher's profile. Based on this profile, the researcher is able to follow other people interested in the same research themes and share their digital libraries and academic productions.

The site Academia.edu is based on social webs. The social network connects the users by means of the content and notes generated by these users. From a perspective of knowledge classification, these researchers share the emerging terminology, which is consolidated within the investigation domains, and provide a classification that reflects the social context and the dynamics of the studies carried out therein.

Relevant terms were selected as they related to the thematic groups available in the Academia.edu platform,

namely: e-Science, cyber-infrastructure, e-Research and Science 2.0. At that stage, 429 profiles of users were found, combined with the proposed themes, with 124 distributed terms. Using the software Visual Anthropac—a freeware program used to assess the cultural domain, we sought to develop a structure for a free list composed of tags created by the users in each of the groups, and identify the projections found in the articulation between the lists obtained (Table 1).

The list of projections clearly referred to the current common belief set forth by the groups upon articulation of the proposed themes; however, oscillations still occur between naming the groups and their tags. There are tag-naming groups concurrently integrating all other groups assessed. From our point of view, that is evidence of a high degree of semantic articulation.

We also sought to identify the thematic interests of eight researchers who were part of the e-Science interest group. In that sense, a list was prepared with the specific terms of each of the researchers, compared to the projec-

TAG	PROJECTION
cloud_computing	0.707
computer_science	0.444
computer_science_education	0.424
computer-based_learning	0.404
and-learning	0.391
digital_culture	0.389
data_mining	0.372
digital_humanities	0.372
distributed_file_systems	0.355
document_genres	0.342
dynamic_load_balancing	0.332
Education	0.306
heterogeneous_distributed_systems	0.279
high_performance_scientific_computing	0.262
information_retrieval	0.214
new_media	0.212
internet_&_society	0.19
open_source_software	0.18
parallel_architectures	0.163
resource_allocation	0.11
social_computing	0.077
software_engineering	0.057
web_2.0	0.024

Table 1. Projection Indicators

tions in the group of tags, with the purpose of identifying similarities and distinctions among them. The result in Figure 2 was achieved.

The observation allowed us to assert that the list obtained represents the concept of e-Science informally agreed upon by the group, as well as to identify the area of interest and of application that prevails among the members. That informal semantic pact is also reflected in the documents shared by the group. Based on the incidence of the terms adopted by the group, it can be observed that, within the assessed group in the Academia.edu website, e-Science is considered a grid of high-performance computers in clouds, focusing on the improvement of research quality in different domains. In that group, 86 terms were found, six of which showed a frequency between 25% and 75%, and indicated that they refer to more consistent terms of the proposed theme. All other 82 terms indicated different approaches of the relevant theme, evidencing a probable semantic or ontological relationship. Among the terms are: digital_humanities, digital_signature, archaeology, archaeological_informatics, authentication, archaeological_graphics_&_illustration, distributed data mining, resource allocation, virtual machines, and information society research.

5.0 Conclusions

The study conducted allowed the observation of a collective effort to settle a semantic interoperability model for

the labeling of contents based on best practices regarding the description of the objects shared in SSIS. The recovery and the negotiation of senses related to the concepts that compose a given discursive formation in collaborative environments worked concurrently as surfaces of emergence, authorities of delimitation and grids of specification surface, allowing more expeditious establishment of signification pacts, both in the labeling and in the information mediation and recovery processes. It was noticed that the formation of emerging online social networks and collaborative environments has been fostering a larger verticality of the treatment of discourse in its generating instance, and strengthening collaboration initiatives among the users, regarding the treatment of emerging concepts (identification, delimitation and specification). In addition, the social semantic information spaces, by allowing the identification of discourse succession and co-existence methods, as well as strengthening of the authors associated with them, represent good perspectives of consolidation of procedures that can represent dynamically the emerging knowledge shared in a network. Such changes in the information and knowledge sharing methods require an ontological and semiotic approach in order to understand the possibilities of social classification and the pacts of signification that arise in the social semantic information spaces. It is believed that this theoretical and technological approach may help to improve semantic tools, classifications, navigation taxonomies and methodologies for the construction of a

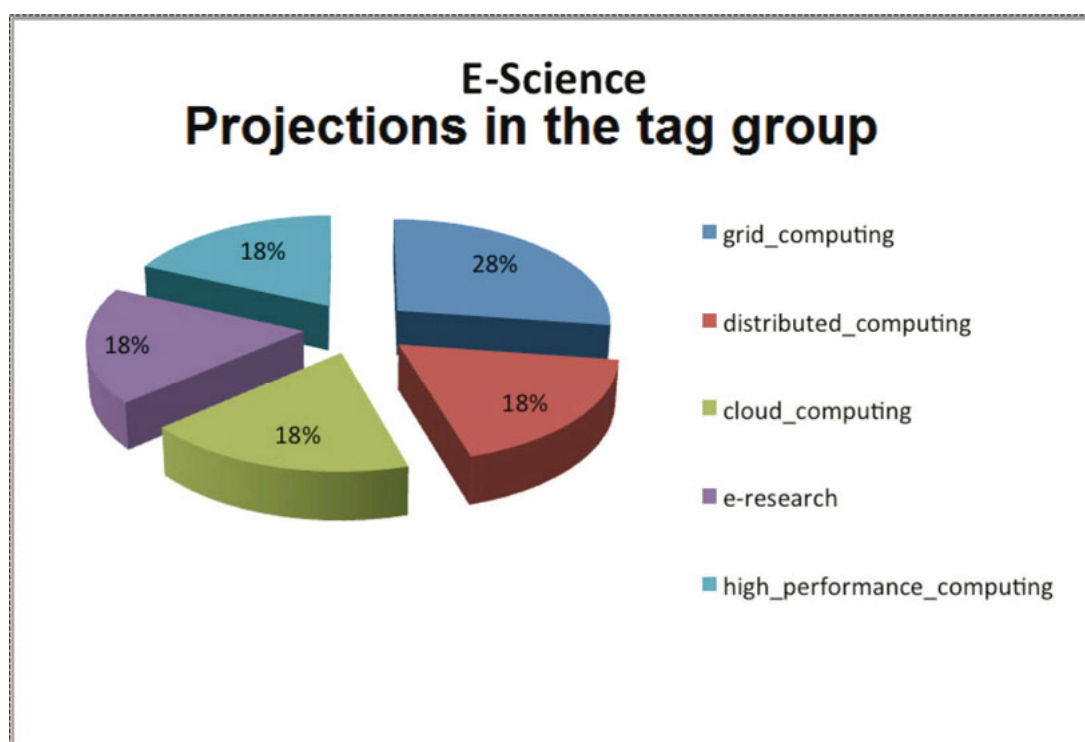


Figure 2

language of indexation in collaborative digital environments.

References

- Domingues, Ivan. 2004. *Epistemologia das ciências humanas. Tomo 1: positivismo e hermenêutica*. São Paulo: Loyola.
- Foskett, D. J. 1962. The Classification Research Group, 1952-1962. *Libri* 12: 127-38.
- Foucault, Michel. 2010. *A arqueologia do saber*. Rio de Janeiro: Forense Universitária.
- Gnoli, Claudio. 2011. Ontological foundations in knowledge organization: the theory of integrative levels applied in citation order. *Scire* 17 no.1: 29-34.
- Goldstein, Jeffrey. 1999. Emergence as a construct: history and issues. *Emergence: complexity and organization* 1: 49-72.
- Hey, Tony and Trefethen, Anne. 2008. E-Science, cyber-infrastructure, and scholarly Communication. In Olson, Gary M, Zimmerman, Ann and Bos, Nathan, eds., *Scientific collaboration on the internet*. Cambridge: MIT Press, pp. 15-31.
- Rivier, Alexis. 1992. Construção de linguagens de indexação: aspectos teóricos. *Revista da Escola de Biblioteconomia da UFMG* 21: 56-99.
- Spiteri, Louise. 1995. The classification research group and the theory of integrative levels. *The Katharine Sharp review* 1. Available <http://web.archive.org/web/20011222083409/alexia.lis.uiuc.edu/review/summer1995/spiteri.html>.
- Spiteri, Louise. 1998. A simplified model for facet analysis. *Canadian journal of information and library science* 23: 1-30.