Formulating knowledge

By Dominique VAUQUIER, Praxademia S.A.S.

Proofreaders and contributors:
Bruno SEZNEC, Thierry BIARD (Ambesas), François-Régis CHAUMARTIN (Proxem)

Translator: Joanne TOWARD

Table of contents

1° Texts: their classification 1

2° Terms: their elucidation using terminology 2

3° Concepts and relations: ontologies and the Web 3

4° Knowledge: its modeling and automation 4

Conclusion 6

Knowledge management comes in various shapes and forms, just as knowledge does itself. Given the stakes that it represents, it is worth listing the different possibilities of capturing, expressing, storing and leveraging knowledge. The object of this paper is not to provide an exhaustive review of the subject or even to present all the possible solutions. It will focus on the expression of knowledge – more or less formal – putting aside the questions linked to how this knowledge is obtained, managed and transmitted. The aim is to present the solutions that are easily accessible to enterprises to express their knowledge. We will demonstrate that they are spread out in a continuum: texts, terms, concepts and models. This continuum allows us to found a progressive approach to knowledge management, highlighting the increasing possibilities for automation.

1° Texts: their classification

Firstly, knowledge appears as texts. When it has still to be formulated, the survey and interview techniques will end up getting it into a text state. Knowledge management that stops at this stage is useful, but it soon comes up against a limitation: these texts must be ordered and fitted into classifications that will enable us to make effective use of them. To this end, we have to build a **classification**, such as the well-known ones used in libraries[1]. Almost always, our culture and – in the case of libraries – the physical constraints of storage lead to a single and hierarchical classification. As our own experience shows us every day, this is not the most natural way of organizing knowledge and preparing its use. Knowledge bubbles away and constantly renews itself; it does not flow easily

[1] For example: the Dewey Decimal Classification, the Universal Decimal Classification, the American Library of Congress Classification and the Freinet Classification.

into a set and definitive mold. At the very least, it is in our interests to adopt not one, but several **taxonomies**[2], each one viewing things from a specific angle.

2° Terms: their elucidation using terminology

From then on, and we are clearly aware of it, things become complicated. How do we identify these taxonomies? What shall we do in order to limit their number? How do we check that they will not start to contradict each other? Etc. Moreover, the texts, beginning with the titles and summaries, will have to be interpreted in order for us to associate the appropriate taxa[3] to them. Clarifying the vocabulary is therefore a precondition for establishing a classification. Here, **terminology**, as a discipline[4], comes to our aid. The terminological analysis is a prerequisite for the subsequent stages which head towards more formalization. In itself, it already produces results that are immediately perceived by the stakeholders. Notably, in the Business Architecture discipline, enterprise terminology, as a key deliverable, is recognized as underpinning the knowledge-based approaches and the convergence between several entities.

Terminology, used in the enterprise, brings the following benefits:

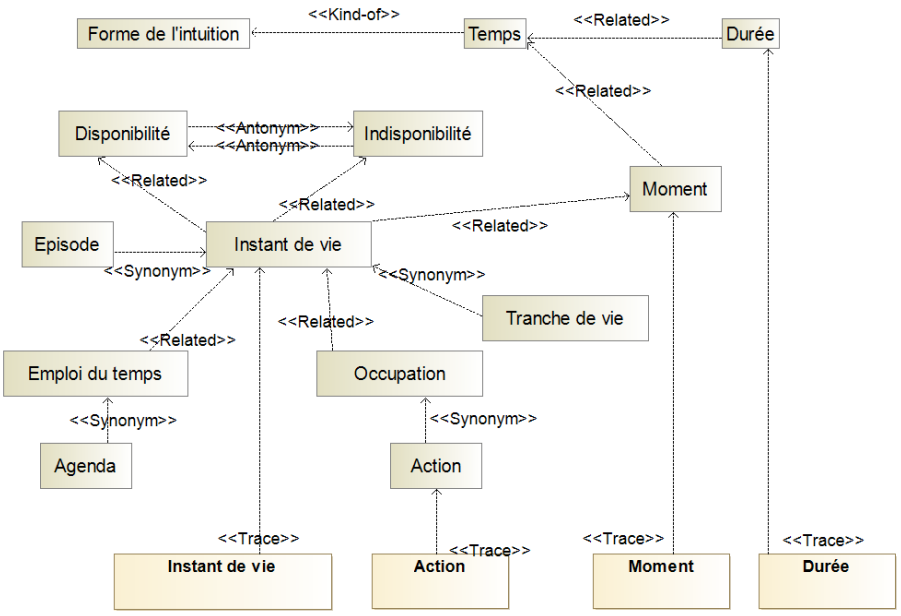
1. It demonstrates “active listening”, through its gathering and analysis of the corpus of texts.
2. It takes the heat out of the quarrels about the terms or definitions[5] by studying the usage and by bringing a sufficiently precise framework to objectify the language.
3. It provides the rules to get to the right definitions, that is to say, definitions that are as simple as possible and which grasp the essence of a concept.
4. It guides us in analyzing the relations between terms, which form the fabric of knowledge[6].

For enterprise terminology, we think of the dictionary form, possibly a “reference” dictionary. If we then add the multiplicity of sources, the study of different usages and, especially, the relations between the terms, we move on to the **thesaurus** form. It is in this form that the enterprise terminology will be the most useful and the most long lasting, in the long term. Today, we have at our disposal efficient tools for developing and managing the thesauri, which can then be published on the enterprise’s intranet, for example.

Terminology helps us to establish better classifications, notably to label our documents. It also helps us to go further in expressing knowledge. Indeed, it organizes the material that will be used from the outset of the modeling.

[2] Taxonomy: classification of elements. SKOS (Simple Knowledge Organization System) is a standard for developing taxonomies and thesauri (see: http://www.w3.org/2004/02/skos/).
[3] Taxon (plural: taxa): an entry in a taxonomy.
[4] In the third meaning as given by the *Grand Robert* French dictionary: “Set of theoretical and practical activities concerning the systems of notions and their designations by means of names organized in a system (terms)”.
[5] Language is a symbolic capital. All capital transforms itself into power. Hence, the impassioned debates that sometimes crop up around terms and their normalization.
[6] Relations between terms have been analyzed and codified for a long time.

Figure 1. Example of a terminological diagram



This diagram was made with a modeling tool that provides a "dictionary" function (Modelio).

It presents terms from the temporality semantic field, linked by typical relations. The rectangles at the bottom refer to semantic classes, the modeling elements that formalize the concepts (within a semantic model).

3° Concepts and relations: ontologies and the Web

Before that, we should tackle another possible stage, that of forming **ontologies**. An ontology (in the IT sense) is a model that links concepts, expressed by terms. Terminology and ontology are dual approaches, fueled by the same theoretical tradition. One feature sets them apart, at least in practice: the importance of tooling in the case of the ontologies[7]. Moreover, the ontologies can be expressed in standardized languages such as RDFS[8] and OWL[9]. These languages form the basis of what is commonly called the semantic Web. The advantage here lies in the sharing of information, which can take the form of complex propositions. A language such as SPARQL[10] enables us to formulate questions and to query one or several RDF triplestores.

In this way, we can see what sets both trends apart:

- Terminology develops for itself, with its terms, its networks of terms (possibly drawn) and its definitions. Once published, it is used like a dictionary, either by clarifying usages or by normalizing vocabulary.
- An ontology is built in the same way, but its representations, perhaps more complex, are reserved for the designers and not designed for the public. Its content, on the contrary, is available and accessible on the Internet or one similar such form, through information systems. The ontology technique sides with the machine-to-machine communication solutions, along with all the solutions that make up the semantic Web.

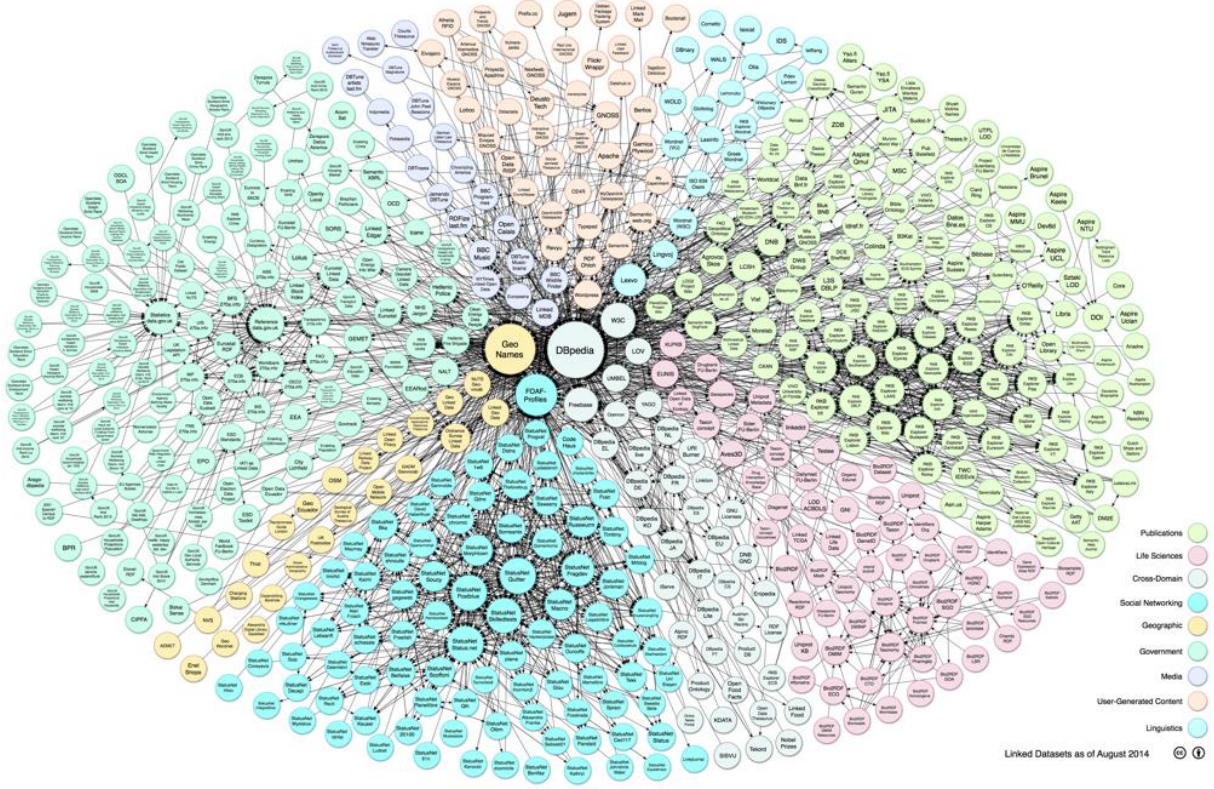
[7] One of the most well-known tools is Protégé (http://protege.stanford.edu).
[8] RDF: Resource Description Framework. RDF enables us to define triples in the form: subject, verb, object. RDFS: RDF Schema. RDFS enables us to represent ontologies.
[9] OWL: Web Ontology Language (an OMG – Object Management Group – standard). Based on RDF, OWL introduces the notion of class, both the concept and the set of instances.
[10] SPARQL stands for "SPARQL Protocol and RDF Query Language". It is a query language adapted to expressions in RDF and used to query the RDF databases.

From a methodological point of view, we would like to recommend both approaches, as they are complementary; one presents the knowledge in an encyclopedic manner, the other in a way that can be exploited from an IT perspective. All the same, this leads to an adjustment between the tools used, if we want to optimize the investment.

In addition, the ontology technique provides the following advantages:

- the reuse of existing ontologies as well as the triplestores available publically[11],
- the interoperability, thanks to standardization (the semantic Web cake which arranges the standards that form the basis of the semantic Web),
- the formal validation capability of an ontology[12].

Figure 2. Linking open data: Linking Open Data cloud diagram



Linking Open Data cloud diagram 2014, by Max Schmachtenberg, Christian Bizer, Anja Jentzsch and Richard Cyganiak. http://lod-cloud.net/

4° Knowledge: its modeling and automation

Finally, one stage remains in our journey towards **formalizing knowledge**. In this last stage, not content with ensuring that the knowledge has been perfectly formulated, with no missing elements and no contradictions, we would like it to be done *formally*, to the extent that it can be understood and reproduced by a machine. This stage relies on models, in particular those produced by semantic

[11] The Linking Open Data cloud diagram (http://lod-cloud.net/) collects the open data sources available.
[12] See, for example, the *reasoned* function in Protégé which deduces the triples following predicate logic "All men are mortals, Socrates is a man, therefore Socrates is mortal."

modeling. To introduce them, let us return briefly to the types of knowledge that we are looking to capture in professional life.

Business knowledge, at work in professional life and without which nothing would get done, takes the following forms:

- knowledge, conceptual knowledge, engrammatic in nature, in the form of concepts and relations between concepts;
- know-how, practical knowledge, which can be expressed by gestures (in manual activities) or operative modes;
- must-dos, imperative knowledge, imposed through procedures and rules;
- duty, deontic knowledge (ethics, values), which constitutes a form of deeply buried knowledge but which is nonetheless necessary to disseminate.

The nature of a knowledge element, that is to say its membership to one of these categories, determines how it is detected and formulated. The Enterprise System Topology[13] sorts these elements out. The table below associates the knowledge categories to the aspect they go under.

Figure 3. Association of knowledge categories to the Enterprise System Aspects

Category	Examples	Enterprise System Aspect
Knowledge	Calculating insurance cover. Client account allocation rules. Performance of a machine. Regulations concerning an activity.	**Semantic aspect**: enterprise aspect made up of the knowledge of the fundamentals of your business. Represented in terms of classes ("business objects"), relations and life cycles.
Know-how	Know how to analyze a client's situation. Know how to build a perfectly vertical wall. Know how to repair a machine.	**Pragmatic aspect**: enterprise aspect concerned with the activities and how they are conducted. Represented in terms of processes, organization, work situations…
Must-dos	Quality procedure. Organization rule defining a power, a role. Security rules…	Between know-how and must-dos, the subtlety is the same as between the description and the instructions. From a knowledge representation viewpoint, the same tools are used.
Duty	To be available for clients. Intervene as soon as possible. Do not do anything that endangers the planet.	**Intentional aspect**, "Values" facet. Here too, we distinguish between the moral reality of the enterprise (observed) and the instructions on how to behave (affirmed).

Here we have, therefore, several models of different types, which will help us to formulate the knowledge. Beforehand, we can ask ourselves whether it is necessary to develop all these models and to cover all these knowledge categories. The answer will depend on the context and on the type of transformation aimed for. Most of the time, the semantic model will be essential as it provides the expression of the concepts, which are the substance for everything that happens in the enterprise.

[13] The Enterprise System Topology is the reference framework proposed by the Praxeme open method. It identifies and articulates seven aspects belonging to all systems and that should be described while respecting their own logic.

The work on the terminology and, possibly on the ontologies, has prepared the semantic modeling. What does the latter bring? The formalism and economy requirements, specific to this discipline, result in several benefits:

- Compared to a terminology or an ontology[14], the semantic model is more compact: it only keeps the concepts that carry useful, characteristic properties; it does not force itself to reuse all the terms[15].
- The semantic model systematically tracks down the generic concepts, often hidden beneath specialized concepts and unnoticed by "business" actors. The genericity of a model is a factor of economy and sharing. In certain cases, it ensures the interoperability with other organizations.
- The model seeks to express the full semantics of the concepts, around three themes: information, transformation and action. It provides itself with categories of representation in the goal of expressing the semantics in such a way that they can be mechanically exploited (operations with a documented signature, state machine…). Different types of relations enable the determinations between concepts to be formalized (inheritance, associations: simple, reified, qualified, N-ary…). The cardinalities of associations greatly contribute to this.
- In so doing, the semantic model ensures a form of verification and completion in the expression of knowledge.
- The semantic model provides the starting point for a revision of the business processes as well as a correct design of the IT solution, by applying "derivation rules" [16].

Conclusion

We have at our disposal several approaches that can express knowledge. They are all useful and we can adopt them simultaneously, either in a gradual approach or for their complementarity. Our main message, in this paper, is that these techniques are organized around a scale that goes from the most intuitive act (classifying documents) to the most demanding (semantic modeling). As you go along this scale, the benefits increase, from a dual point of view:

- the certainty of having expressed *all* the knowledge, in a correct way, that can be verified and even activated;
- the ability to automate all or part of this knowledge.

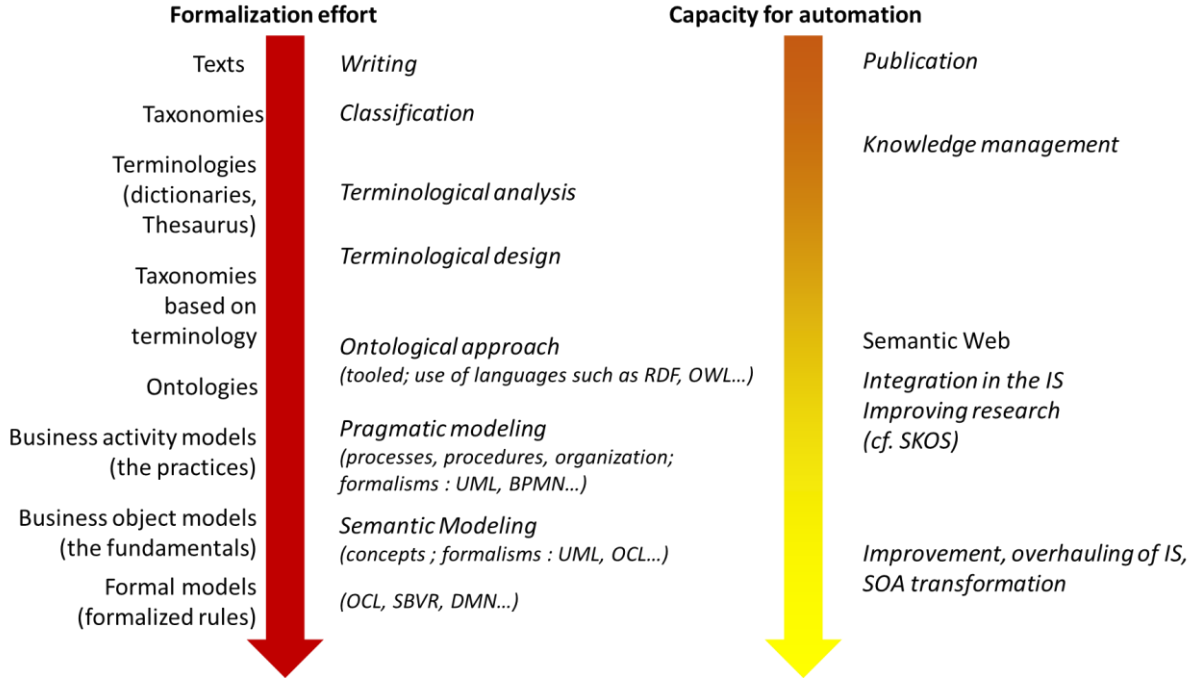
The diagram hereafter summarizes this position. The Praxeme enterprise transformation methodology seeks to turn to good account the different disciplines, some of which are confined to the academic sphere. Its vocation is to bring them into the enterprise to help with its transformation. To this end, it uses the different disciplines to draw up procedures that make up as many manuals to guide practices. As its ambition is to cover all aspects of the enterprise, it can indicate how these procedures are positioned and linked together in a controlled transformation effort.

[14] An ontology can easily mix classes and instances at the same level – which would cause a semantic modeler to shudder. Semantic modelers use classes and instances, but rigorously separate them (impossible, for example, to represent them in the same diagram).

[15] The modeler applies the criterion of Ockham's razor, systematically eliminating the elements that we can do without in the model, without reducing its expressive power.

[16] The mechanical exploitation of knowledge is made possible by the scope of the method that covers all aspects of the enterprise.

Figure 4. Gradation of the effort in formalizing knowledge and level of expected results



Other solutions or approaches concern how knowledge is formulated. We can mention, in particular:

- natural language processing (NLP) which enables us to exploit corpora of texts or messages to extract information from them (right up to the feelings expressed in letters)[17],
- big data, thanks to which new knowledge is emerging through advanced analysis techniques (including machine learning).

So, we can see that enterprises, which recognize the stakes, have a wide range of solutions at their disposal for expressing knowledge. They have to decide upon their strategy by combining these solutions and by balancing the formalization effort, related to the consequences in terms of control and automation.

Dominique.vauquier@praxeme.org

Consulting and training dedicated to the development and implementation of the open method

http://www.praxademia.com

[17] See the offer from the company Proxem and, more generally, the natural language processing solutions.