Collaboration and the Knowledge Economy: Issues, Applications, Case Studies
Paul Cunningham and Miriam Cunningham (Eds)
IOS Press, 2008 Amsterdam
ISBN 978–1–58603–924-0

Semantic Technologies for European Chambers of Commerce Cooperation

Michele MISSIKOFF, Francesco TAGLINO
LEKS IASI-CNR, Viale Manzoni 30, Rome, 00185, Italy
Tel: +39 06 7716433, Email: missikoff@iasi.cnr.it, taglino@iasi.cnr.it

Abstract: This paper presents the semantics-based solution developed and applied in the LD-CAST project, which aims to support the cooperation of Chambers of Commerce in an European network. The solution is based on three main elements: (i) a Weighted Reference Ontology (WRO), an ontology where each concept is enriched by a weight expressing the selectivity degree of a concept that annotates a resource; (ii) a semantic annotation technique based on Ontology Feature Vectors (OFVs), where business process (BP) activities and concrete services (CSs) are annotated with a set of concepts from the WRO; (iii) a semantic similarity search method which allows to identify the CSs suitable to perform the BP activities, in order to make the BP executable.

Keywords: Reference Ontology, Semantic annotation, Semantic similarity, Business Process

1. Introduction

In this paper we present the key elements of the semantic solutions adopted in the LD-CAST platform aimed at supporting the cooperation of Chambers of Commerce (CoC) within a European network[i].

One of the main objectives of the European CoC is to assist SMEs when they start a new cross-border business or when they face problems in conducting such a business. To this end, in the LD-CAST Project, a semantic cooperation platform, based on a reference ontology, has been implemented. Such a platform provides flexible services to support the different phases of a cross-border cooperation, overcoming many of the hindrance factors (e.g., different business models and information systems).

Semantic technologies represent a central part of the LD-CAST Platform. They cope, on the socio-organizational level, with the diversity of organizational models characterizing the different Chambers of Commerce; on the implementation level, with the diversity of specifications and diverse structure of business processes and activities composing them. Furthermore, the platform supports the late binding of concrete services implementing specific operations, offered by different service providers, that will be selected at runtime through the applications of semantics-based similarity techniques.

The platform is based on a phased approach organized into setup-time and run-time (see Figure 1) [3]. The first requirement for the setup is the availability of a Weighted Reference Ontology (WRO) where each concept is enriched by a weight, expressing the selectivity degree of each concept in characterizing a resource. In LD-CAST, the Reference Ontology has been built by using the Athos Ontology management system [2]. The WRO is used to semantically annotate the available resources: the business processes, the activities composing them, and the concrete services (executable service implementations). Each annotation is represented by an Ontology Feature Vector (OFV), which associates a set of concepts to a resource. The semantic annotation allows a semantic alignment to be established between the activities composing a BP, which identify what has to be done

(e.g., RequestingForLegalVerification), and the concrete services (CS), the actual available implementations of such activities. BPs are modelled by the CoC by using the ADOeGov modelling tool [10]. Since the concrete services are made available by different service providers, their descriptions are not homogeneous and, in principle, not aligned with the activities definition specified in the BP. Semantic annotation allows a semantic bridge between these two artefacts (i.e., BPs and CSs) to be built.

Having achieved the ontology building and the semantic annotation, which represent the steps of the setup phase, at run-time, a semantic search, guided by the WRO and the semantic annotations, is used for selecting the concrete services suitable for executing the activities of the business process activated by a user. In order to accomplish the binding of concrete services to BP activities (Business Process completion), the role of the ontology is crucial and it is used at two different levels:

- *explicitly*, when the ontology is directly accessed by the Semantic Search and Discovery Engine (SSDE) for querying its semantic content (e.g., navigating the Specialization hierarchy);
- *implicitly*, when the semantic annotation expressions, built in terms of the reference ontology, are used by the Semantic Search and Discovery subsystem.

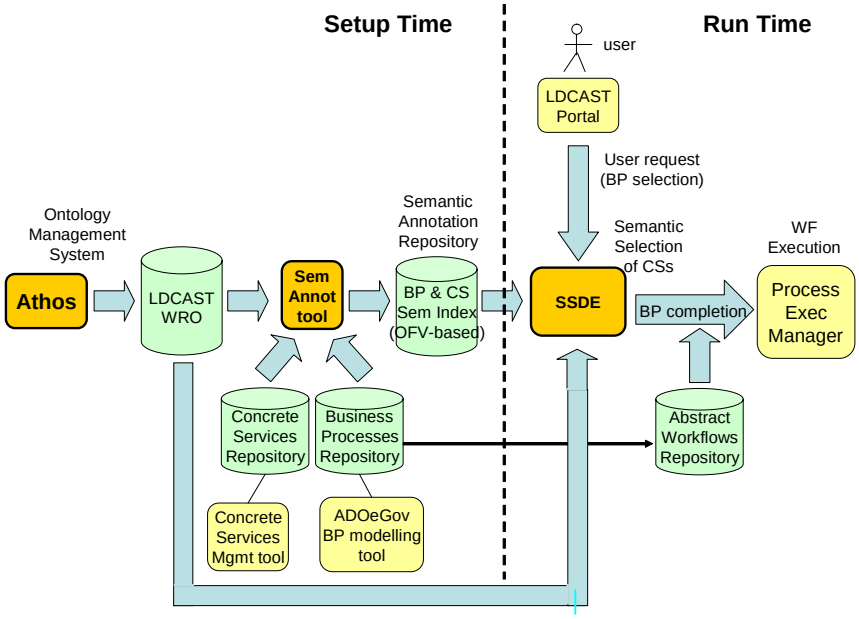

Figure 1. The LD-CAST Semantic streamline

The rest of the paper is organized as follows. Section 2 presents related works on ontology-based solutions in the eGovernment domain. Section 3 presents a methodological overview of the proposed approach. In Sections 4, 5, 6, the characterizing aspects of the solution, that are the WRO, the OFV-based annotation and the semantic similarity search, are described. Section 7 presents the application context in which the work has been carried out. Finally, the Section 8 presents the conclusions.

2. Related Works

The e-government scenario is a promising application field for ontologies, since legislative knowledge is by nature formal to a large extent and its definition is shared by many stakeholders. In fact, there are several initiatives addressing eGovernment problems by adopting semantics-based solutions. Among them we wish to cite the following European projects: Ontogov [1] developed a platform to facilitate the composition, reconfiguration and evolution of e-government services. Qualeg [14] developed methodologies and

software solutions enabling a local government to manage its policy lifecycles. Terregov [17] addressed the issue of interoperability of e-Government services for local and regional governments. SmartGov [18] developed a knowledge-based platform for assisting public sector employees to generate on-line transaction services. ICTE-PAN [12] developed methodology for modelling and redesigning PA operations, and providing supporting tools for transforming these models into design specifications for e-Government collaboration environments and portals. e-POWER [5] employed knowledge modeling techniques for inferences like consistency check, harmonisation or consistency enforcement in legislation.

With respect to other projects, LD-CAST is characterized by two key solutions. The first is the integrated, semantically enriched knowledge repository that spans from BPs to laws and regulations. The second is the composition of business process workflows achieved dynamically by using semantic search and discovery solutions, which are the focus of this paper.

3. Methodological overview

The core of the proposed semantic solution is represented by the three following elements.

- The weighting of the concepts in the ontology, yielding a Weighted Reference Ontology (WRO). Intuitively, the weight of a concept indicates its selectivity degree in characterizing a resource.
- The Ontology Feature Vector (OFV) annotation method. An Ontology Feature Vector is a set of concepts from the ontology, each of which concurring to characterize the annotated digital resource (e.g., business processes or concrete services). Such an annotation method is based on the availability of a Weighted Reference Ontology. The Annotation of BPs yields the so called *Semantic BP*, where each activity composing the process is replaced by an OFV.
- The semantic search and retrieval (SSR) technique that is applied to perform the concrete services discovery and selection, in order to bind concrete services to business process activities. The SSR method is based on semantic similarity reasoning techniques. For each activity composing the requested BP, the Semantic Search and Discovery Engine retrieves a set of candidates services (SCS), i.e., CSs that exhibit a similarity degree above a given threshold. From the semantic point of view, all the CSs in a SCS are considered equivalent. The final selection to identify the actual CSs to be bound to the BP is based on quality of service criteria expressed by the user.

Figure 2 depicts the entire approach. The next three sections are dedicated to the description of how the SCS for a certain business activity is identified, in order to create the Generic Workflow for a given BP. The Generic Workflow has the same topology as the corresponding BP, but each activity is substituted by the corresponding SCS.

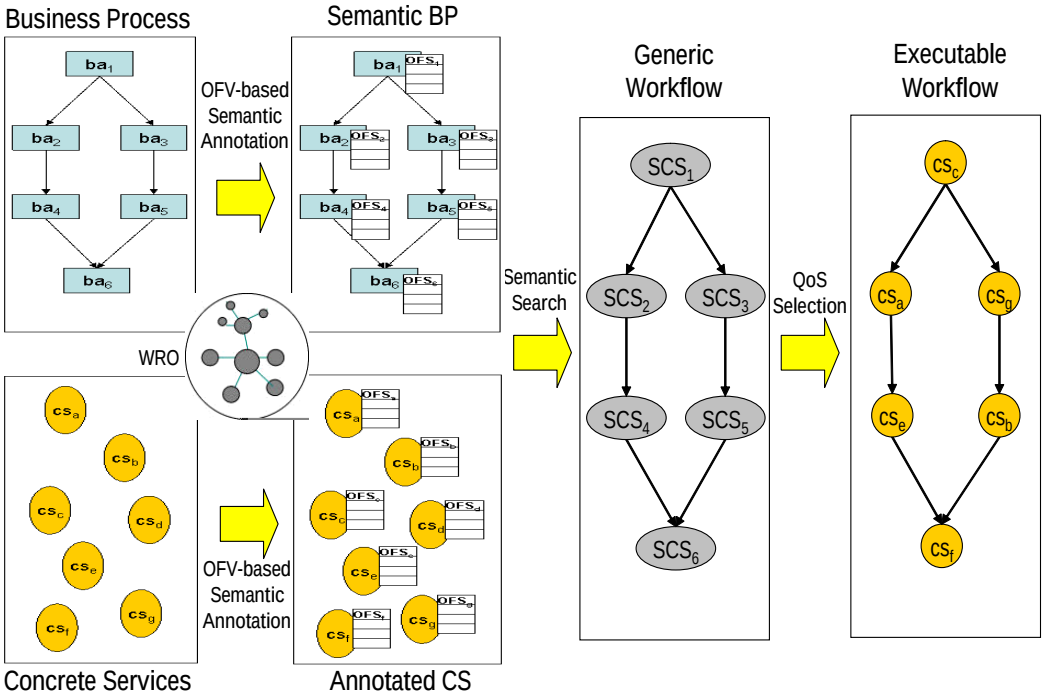

Figure 2 – From Business Process to Executable WF

4. The Weighted Reference Ontology

At the basis of the proposed method there is a Weighted Reference Ontology (WRO). Firstly we wish to recall the definition of an ontology, taken from OMG Ontology Definition Metamodel [4]:

> "An ontology defines the common terms and concepts (meaning) used to describe and represent an area of knowledge. An ontology can range in expressivity from a Taxonomy (knowledge with minimal hierarchy or a parent/child structure), to a Thesaurus (words and synonyms), to a Conceptual Model (with more complex knowledge), to a Logical Theory (with very rich, complex, consistent and meaningful knowledge)."

Our view of the ontology is restricted to a taxonomy of concepts, i.e., the ISA hierarchy. This is a traditional view, but in our proposal the innovative part is represented by the fact that each concept in the ontology is associated with a weight that represents its selectivity degree in characterizing a resource. In essence, in accordance with the Information Theory [15, 16], we start from the measure of the probability that a concept characterizes a resource. It means that the more the concept is generic the higher is its probability. Conversely, a low weight corresponds to a high selectivity, and therefore its characterizing power will be more relevant. In accordance with the Information Theory a concept weight will be used to determine the (relative) information content of each concept.

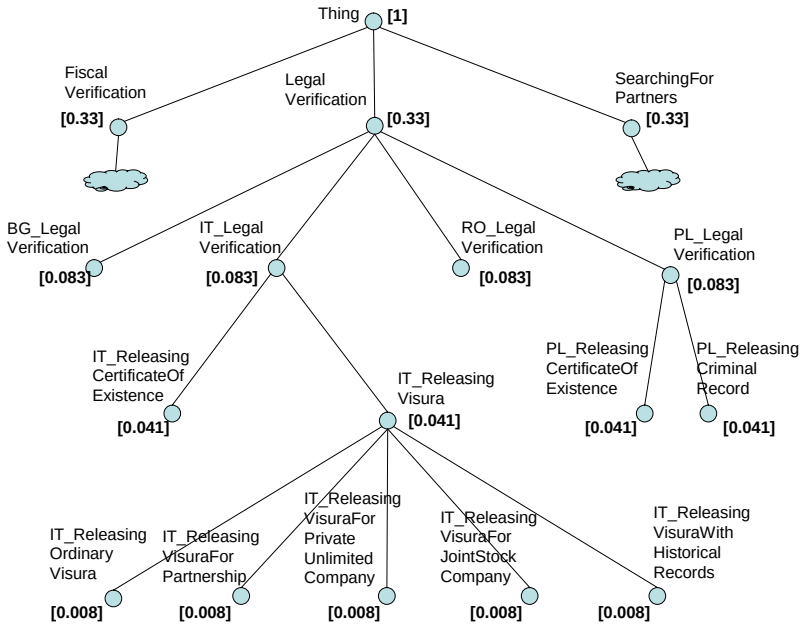

Figure 3. An example of Weighted Reference Ontology

To associate a probability to each concept, we adopted a uniform probability distribution, based on a top down approach that traverses the ISA hierarchy. In order to have a unique top concept in the ISA hierarchy, the most general concept Thing is represented as the root of the hierarchy. Whatever is the ontology, and consequently its ISA hierarchy, the probability associated with Thing, p(Thing), will always be equal to 1. This is justified by the fact that any resource can be annotated with such a top concept, even if such an annotation is not meaningful (i.e., it has zero selectivity).

For any other concept *c*, p(c) is equal to the probability of the father of c, divided by the number of the children of the father of c (i.e., the fan-out). In Figure 3, an excerpt of the ISA hierarchy, from the WRO built in the LD-CAST project, is reported. The corresponding probability is associated with each concept. For instance, considering the concept IT_LegalVerification:

$$p(IT\_LegalVerification)=0.083,$$

since p(LegalVerification)=0.33 and LegalVerification has 4 children.

Since the probability decreases from the top to the bottom of the ISA hierarchy and whatever is *c*, $0 \leq p(c) \leq 1$, we have that $\log(p(c))$, which represents the information content carried by the concept *c*, increases from the bottom to the top. Finally, $\log(p(Thing)) = 0$ because Thing is the top of the specialization hierarchy and therefore has a null selectivity degree.

5. OFV annotation of Business Processes and Concrete Services

As anticipated, the semantic annotation method proposed in LD-CAST is based on Ontology Feature Vectors (OFVs). An OFV is a set of concepts from the ontology, each concurring to characterize the annotated digital resource.

The terminology has been borrowed from the Feature Vector theory[ii], as a synthetic representation of some object. The Feature Space from which the features are extracted is represented by the WRO.

OFV-based semantic annotation method is here used to annotate Business Processes, and in particular the activities composing them, and the Concrete Services. An example of annotation of BP and CS follows.

In the proposed example, the Business Process is: CheckingTheStatusOfItalianCompany (see Figure 4).

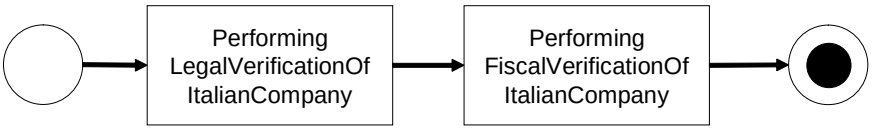

Figure 4. Checking the Status of an Italian Company Business Process

This process aims at checking the legal and fiscal status of an Italian Company. To this end, it is composed by two activities:

PerformingLegalVerificationOfItalianCompany and

PerformingFiscalVerificationOfItalianCompany

Annotating the BP means to annotate the two activities composing it. According to the WRO in Figure 3, the annotations, that are the corresponding OFVs, are the following:

OFV(PerformingLegalVerificationOfItalianCompany) = [IT_LegalVerification]

OFV(PerformingFiscalVerificationOfItalianCompany) = [IT_FiscalVerification]

Concerning the Concrete Services, it is here reported the annotation of three services (CS1, CS2 and CS3) concerning the Legal Verification. The corresponding annotation, which better characterizes their functionalities follows:

OFV(CS1) = [BG_ReleasingVisura, PL_LegalVerification, RO_LegalVerification]

OFV(CS2) = [IT_ReleasingCertificateOfExistence, IT_ReleasingVisuraForPartnership]

OFV(CS3)= [IT_ReleasingOrdinaryVisura, IT_ReleasingVisuraWithHistoricalRecords]

6. Semantic search for BP workflow completion

Semantic Search of Concrete Services is performed by applying a semantic similarity search method, which needs to access the Reference Ontology and the Semantic Annotation of BPs and CSs performed in the setup phase. The aim is to discover the available CSs suitable for performing the activities composing the BP selected by the user. To this end, the OFVs associated to the BP activities will represent the search criteria, here called Request Feature Vectors (RFVs).

Considering the example in the previous section, we will consider the searching for the CSs performing the PerformingLegalVerificationOfItalianCompany business activity. In this case, RFV = [IT_LegalVerification].

Starting from the RFV, the semantic search has to evaluate the semantic similarity degree between the RFV and the OFV associated to each CS. All the CSs whose OFV has a similarity degree not less than a given threshold, are put in the Set of Candidate Services. The search and retrieval method is based on semantic similarity reasoning techniques, the semsim function and the consim function. The semsim, evaluates the similarity between two OFVs, and it needs the consim, which evaluates the similarity between a pair of concepts.

Given a WRO, the notion of consim relies on the probabilistic approach defined by Lin [11], which is based on the notion of information content. Formally, we have

$$\text{consim}(c_1,c_2)=2 \log p(\text{lub}(c_1,c_2)) / (\log p(c_1)+\log p(c_2))b$$

where -log p(c) is the information content carried by the concept c, lub(c_1,c_2) represents the least upper bound of c_1 and c_2 in the WRO (i.e., the nearest ancestor), and -log p(lub(c_1,c_2)) represents the maximum information content shared by the two concepts.

The semsim requires the evaluation of the consim between each concept in the RFV and each concept in the OFV annotating any CS.

Now, the computation of the semantic similarity semsim, between the RFV and an OFV, is inspired by a method based on the maximum weighted bipartite matching problem

in bipartite graphs [6,7,9]. The method adopted to solve this problem is based on the so called Hungarian Algorithm [13].
We recall to [8] for further details on *consim* and *semsim* functions.
According to the above example, the results are here reported:

semsim(RFV, OFV(CS1)) = 0
semsim(RFV, OFV(CS2)) = 0.87
semsim(RFV, OFV(CS3)) = 0.*68*

Assuming that the threshold has been fixed to 0.8, CS2 will be selected as a candidate service and, it will be the unique CS in the Set of Candidate Services for the IT_LegalVerification activity.

7. The LD-CAST application context and lessons learnt

In the LD-CAST project three main Business Processes have been addressed: Legal Verification, Fiscal Verification and Search for Partners. All the three cases have been designed for the four countries whose Chambers of Commerce are involved in the project: Bulgaria, Italy, Poland, and Romania. Accordingly, the WRO has been built and it is currently composed by more than 300 concepts. The platform has been tested in several experiments where the Chambers of Commerce, partners of the LD-CAST project, have involved possible customers and service providers. The tests aimed at two different purposes, on the one hand to assess the usability of the platform and on the other to let possible service providers to know the platform itself, in order to encourage the development of concrete services as eServices.

Services currently offered by the Chambers of Commerce are mainly requested by the interested companies in traditional ways, such as personal visits (desk service), phone calls, or e-mails. In limited cases (mostly in Italy), cross-border services are offered online on the providers websites. Local Chambers of Commerce analysed have expressed willingness to improve the quality of their offer providing some of their services online. This shall lower the cost of services provision, speed up the execution of requests, and finally allowing new customers to be reached. Today, especially in Eastern EU countries, the e-government is a key element of the transition from industrial to information society and serves to accelerate the European integration process. We believe that the LD-CAST semantic platform will contribute to demonstrate the feasibility and validity of eGovernment initiatives.

Finally, one of the relevant reusable results, part of the LD-CAST semantic platform, is the Weighted Reference Ontology. We started to define a shared and agreed glossary of terms to be used by all partners addressing the same issues. Coming from different European countries, with different cultures and legislations, project partners had initially a hard time in coming to common definition of business terms. This is a very fundamental task to be achieved in setting up a system as LD-CAST that addresses cross-border problem areas. The result of this work will be further improved after the project conclusion, and made available on a consultancy basis to organisations dealing with cross-border cooperation among enterprises, especially SMEs.

8. Conclusions

This paper presented ontology-based solutions providing semantic services aimed at supporting the cooperation of European Chambers of Commerce. The solution is characterized by three main aspects: the Weighted Reference Ontology (WRO), where each concept is enriched by a weight corresponding to the probability that it characterizes a given resource; the OFV-based semantic annotation, where an OFV is associated to the activities of business processes and to concrete services; the semantic search of CSs for the

BP Workflow completion guided by a semantic similarity matchmaking technique based on the semsim and consim functions.

Furthermore, the same semantic approach based on the *semsim* and *consim* functions has been applied in a tourism domain experiment where it has been compared with human judgement results [8] with encouraging results.

References

[1] D. Apostolou, L. Stojanovic, T. Pariente Lobo, and B. Thoenssen, *Towards a Semantically-Driven Software Engineering Environment for eGovernment* - TCGOV 2005, LNAI 3416, pp. 157 –168, 2005.

[2] ATHENA, D.A3.1. SoA on Ontologies and the Ontology Authoring and Management System, with Ontology Modelling Language. ATHENA IP, deliverable, 2005.

[3] De Nicola A., Missikoff M., Taglino F.. A Semantic Cooperation Platform for the Interoperability of European Chambers of Commerce. In the Proceedings of the eChallenges e-2007 Conference, The Hague, The Netherlands, 24th-26th October, 2007.

[4] DSTC, IBM, Sandpiper Software; Ontology Definition Metamodel; Revised submission to OMG, 10 January 2005; http://www.omg.org/docs/ad/05-01-01.pdf

[5] E-POWER Project Home Page, Available at http://www.lri.jur.uva.nl/~epower/

[6] Formica A.; Concept similarity by evaluating Information Contents and Feature Vectors: a combined approach; To appear in Communications of the ACM (CACM), 2007.

[7] Formica A., Missikoff M.; Concept Similarity in SymOntos: an Enterprise Ontology Management Tool, Computer Journal 45(6) (2002) 583-594.

[8] Formica A., Missikoff M., Pourabbas E., Taglino F.. An Ontology-based Similarity Method for Semantic Search. To appear in proceedings of IST-Africa Conference 7-9 May. Maputo, Namibia

[9] Galil Z.; Efficient algorithms for finding maximum matching in graphs, ACM Computing Surveys 18(1) (1986) 23-38.

[10] LD-CAST, D3.1. Common knowledge base specifications. Deliverable 2006

[11] Lin D.; An Information-Theoretic Definition of Similarity, in: J. W. Shavlik (Ed.), Proc. of 15th the International Conference on Machine Learning, Madison, Wisconsin, USA, Morgan Kaufmann, 1998, 296-304.

[12] E. Loukis, S. Kokolakis. Computer Supported Collaboration in the Public Sector: The ICTE-PAN Project. LNCS 2739. Berlin, Springer-Verlag, pp.181-186, 2003.

[13] Munkres J.; "Algorithms for the Assignment and Transportation Problems", Journal of the Society of Industrial and Applied Mathematics, 5(1):32-38, 1957 March.

[14] QUALEG, D3.4 *Ontologies & Semantic Engineering in QUALEG*. QUALEG STREP in 6FP–IST. Deliverable D3.4 September 2006.

[15] Ross S.; A First Course in Probability; Macmillan; 1976.

[16] Shannon C.E.; A Mathematical Theory of Communication; The Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656, July, October, 1948.

[17] Terregov, *Semantic based Support to Civil Servants.* Terregov STREP in 6FP–IST. Del. D3.3 June 2006.

[18] C. Vassilakis, G. Lepouras, S. Rouvas, P. Georgiadis. *Integrating e-Government Public Transactional Services in the Public Authority Workflow*. Electronic Government Journal, vol. 1, 2003

---

[i] This work has been partially supported by the LD-CAST European Project, FP6-2004-IST-4 No. 26919

[ii] A feature vector is an n-dimensional vector of numerical features that represent some object... [it is obtained as a] reduced representation set of features (also named features vector). Transforming the input data into the set of features is called features extraction