ARISTOTELE: A Semantic-driven Platform for Enterprise Management

Pierluigi Del Nostro, Francesco Orciuoli, Stefano Paolozzi, Pierluigi Ritrovato, Daniele Toti
Research Center in Pure and Applied Mathematics (CRMPA), MOMA S.p.A.
Salerno, Italy
{delnostro, paolozzi, toti}@crmpa.it, {orciuoli, ritrovato}@momanet.it

Abstract—We present the architecture of the ARISTOTELE platform, a semantic-based collaborative system for managing enterprises and organizations, specifically designed to include a variety of features. These range from top-level functionalities like managing enterprise processes and building innovation, to finer-grained tasks like customized support for the daily activities of workers, including the creation and execution of personalized learning activities via an adaptive/non-adaptive strategy and the acquisition and usage of collaborative knowledge by the members of the organization. Here, we motivate ARISTOTELE’s compliance with a known Enterprise Architecture framework and describe the design methodology behind the platform and its building blocks, beginning with its data layer and then proceeding to detailing its core services and the higher-level tools built on top of them.

Keywords—Semantic Enterprise architecture; knowledge models; Human Resource Management; Learning; Training; Collaboration; Knowledge Management; Innovation

I. INTRODUCTION

Information is a critical asset for organizations and their members, and being able to effectively harness it and turn it into knowledge has become more and more important a factor for the success or failure of the organizations themselves.

A wide range of solutions and products have been proposed so far and/or are available in the industry to support enterprises and their workers in managing processes and knowledge, but they are usually either too general or fulfill specific, often standardized tasks. Over the course of the last ten years, the advancements in semantic-based research and technologies have progressively brought about new possibilities, as well as challenges, for software systems which might be able to perform a leap forward in terms of knowledge acquisition and management, in order help an organization dramatically increase its efficiency and productivity.

In this regard, ARISTOTELE places itself as an innovative system, being a full-fledged architectural solution which includes several advanced features for managing enterprise-related knowledge in a single, cohesive solution.

Several challenging and interdisciplinary research themes are strictly related to the design and development of such a platform, such as didactic modeling, collaborative innovation and working environments, knowledge management and discovery, collective intelligence, adaptive and non-adaptive learning strategies and so forth.

In fact, the ARISTOTELE platform has been designed as an integrated set of online tools with the purpose of providing workers, managers, trainers and other stakeholders involved in an organization with functionalities and resources in order to support and enhance the growth of their competences and creativity, by assisting them in the process of acquiring, handling and sharing new information and knowledge among them.

The architecture of the ARISTOTELE platform allows for the semantic representation of the information within the adopting organization, enabling open innovation through the Linked Data paradigm, and thus allowing the organization to take advantage of resources that are outside its boundaries, as well as and contributing to the expansion of the publicly available knowledge on the Web.

The functionalities of the platform are designed on top of an appropriate set of methodologies and techniques developed to address those challenging problems, via a number of tools made available to the users in order to take full advantage of such methodologies.

A set of models — the ARISTOTELE Models — is employed for the representation of the domains of interest. These models are in fact used for the ontological specification of semantic aspects of interest, transversal to a wide number of organizations, allowing for the specification of taxonomies for classifying information in each specific context.

Even though each tool of the platform tackles challenges and problems of a specific key area, the added value of ARISTOTELE lies in its capability of achieving a synergy among its underlying methodologies, by seamlessly integrating all of the various tools into a common platform and thus enhancing their mutual cooperation.

The purpose of this paper is to provide a description of the methodologies and techniques exploited to define ARISTOTELE’s software architecture, as well as an overview of its most relevant components. Such an architecture is defined as the set of elements needed to reason about the system, including software modules, their mutual relationships, and the properties of both, and it is based upon a well-established framework for Enterprise architectures.
The paper is structured as follows. In Section II, we list some related work and a handful of well-known frameworks and solutions for defining Enterprise architectures. In Section III we motivate the reasons behind our chosen compliance with one of those frameworks, underlining key aspects of ARISTOTELE in accordance with the framework’s recommendations. In Section IV, we provide a more detailed description of ARISTOTELE’s architecture, by listing its tools, services and data layer. And finally, in Section V we draw our conclusions.

II. BACKGROUND AND RELATED WORK

Despite the fact that ARISTOTELE is a research project, it possesses aspects that can be framed within the context of Enterprise Architectures: several widespread frameworks for the definition of Enterprise Architectures have been thus analyzed, in order to assess how ARISTOTELE fared in comparison by evaluating its overall level of compliance with them. The frameworks taken into account are described below.

- TOGAF[18]. This is an Open Group industry standard comprising a framework with a methodology for the description of the state of the art and possible future states of an organization, with the goal of improving business efficiency and IT support. TOGAF’s documentation is quite wide and covers all the aspects of an organization: business processes, hardware/software infrastructure and data architecture.

- Zachman[19]. This framework provides a means of classifying an organization’s architecture from different point of views, by considering the target perspective (planner, designer, builder etc.) and the focus of the description (functions, data, network etc.).

- Method for an Integrated Knowledge Environment (MIKE2.0) [14]. MIKE2.0 is an open-source framework for Enterprise Architectures with a specific focus on Information Management. Basically, this is a sort of toolbox for defining an Enterprise Architecture, with the purpose of supporting the architect by providing him/her with a general methodology, a reference architecture, a governance model and a number of assets for supporting specific activities. This framework is currently being built collaboratively by many contributors, by in turn applying the MIKE2.0 methodology for the definition of the framework itself. Aside the general methodological aspects, MIKE2.0 provides a set of solutions as follows:
  - Core Solutions, satisfying the needs of a number of organizations (Enterprise Data Management, Business Intelligence, Information Asset Management etc.)
  - Composite Solutions, which are suited for specific contexts (Agile Information Development, Enterprise 2.0, Information Sharing, Semantic Enterprise)

All of the aforementioned frameworks aim at describing the state of an enterprise organization in terms of all the possible dimensions of interest (people, business processes, strategies, IT support, data, network, etc.), while at the same time providing the means to design solutions for the evolution of the organization. Despite Zachman’s is called a framework, it can be considered more as a way of organizing document artifacts that describe an architecture rather than a real framework with a well-defined methodology. Both TOGAF and MIKE2.0 frameworks, instead, propose iterative methodologies built on a set of phases that are cyclic after an initial priming phase. In this regard, the TOGAF framework has a considerable degree of generality and covers aspects that are beyond the scope of the ARISTOTELE project. MIKE2.0, on the other hand, focuses on information management and provides specific guidelines for solutions with respect to knowledge management, social collaboration and semantic management of information.

According to these considerations, we ultimately selected MIKE2.0 as the most suitable framework to be used within the context of the ARISTOTELE platform. Further details regarding the methodology used to bring about the architecture of the ARISTOTELE platform and its compliance with MIKE2.0 will be described in Section III.

Due to the innovative nature of the ARISTOTELE platform, there are few full-fledged architectural solutions to be actually used for an effective comparison. On the other hand, a certain number of research and industrial proposals feature a partial subset of ARISTOTELE’s capabilities, being each of those proposals focused on specific aspects of enterprise-related knowledge.

For instance, PROLIX[17] features a process-oriented learning approach built with a flexible architecture based on the SOA (Service-Oriented Architecture) paradigm. Its core objectives revolve around filling competence gaps for workers, tailoring learning material to their needs (by using a proper methodology for matching required competences with workers’ profiles) and providing ad-hoc fruition of such a material. As it stands, PROLIX focuses on coupling business processes with learning processes in corporate environments and on providing competence-based decision support for those processes, whereas it does not specifically take into account social relationships among workers within an enterprise.

MATURE[16], instead, is made up of a Conceptual model of enterprise knowledge, an employee-level collaborative platform to be integrated into the working environment of learners, and an organization-level platform to encourage innovation and collaborative activities.

As far as specific tools are concerned, Cogito Answers[9] is a semantic search platform with a Natural Language Interface (NLI) providing users with access to unstructured
III. A METHODOLOGY FOR A SEMANTIC ENTERPRISE ARCHITECTURE

As we said in Section II, we have based the design methodology for the ARISTOTELE platform and its under-lying architectural choices upon the MIKE2.0 framework. Specifically, the Semantic Enterprise solution described in MIKE2.0 promotes the adoption of Semantic Web standards (RDF, OWL, SPARQL) for data interoperability and integration, in order for them to be effectively used to fulfill the Linked Data paradigm. MIKE2.0’s suggestions in this regard are related to the possibility of semantically indexing either structured or unstructured information assets, as well as to the clean separation within semantic data between models (concepts and their relationships) and their corresponding instances. Furthermore, MIKE2.0 underlines the need of spreading collaborative technologies for the construction of enterprise social networks, striving towards user-driven content and content sharing. It also encourages the use of taxonomies to define the structure for interconnecting disparate contents and classify information according to the user’s choices. In addition to this, it emphasizes the activity of designing user interfaces, deeming it a critical factor for the successful adoption of a determined system.

By keeping those guidelines in mind, it is fairly simple to assess ARISTOTELE’s compliance with the MIKE2.0 framework. Actually, ARISTOTELE widely exploits semantic web technologies according to the standards mentioned in such a framework. As a matter of fact, ARISTOTELE is able to produce a semantic network so that it might enable interoperability and integration of information within the adopting organization, by resorting to Knowledge Building techniques in order to conceptualize disparate contents and integrate them into a semantic repository. Similarly, when information is created by the ARISTOTELE tools, such information gets simultaneously interlinked and embedded into its semantic network, which is conceived with a clean separation between models (containing domain concepts and relationships) and instances. This way, the Linked Data paradigm is brought to life within the context of the organization, by linking all of its information assets together and allowing them to be shared outside the organization itself.

Also, in ARISTOTELE there are a number of social features that comply with the indications provided by MIKE2.0. Some of those features were already available out-of-the-box in the commercial systems used within the platform (listed in Section IV), while others have been specifically designed for ARISTOTELE.

In terms of software architecture, the ARISTOTELE platform has been designed by relying on the principles of Service Oriented Architecture (SOA), which is a design approach where business functionalities are brought to fruition and offered by means of services: these are standardizable, reusable, interoperable and composable software components, whose interfaces can be published and discovered to enable invocation by heterogeneous clients.

In the next section, we will delve deeper into the architectural structure of ARISTOTELE, by providing a description of its building elements along with their corresponding placement within the whole platform.

IV. ARCHITECTURE OF THE ARISTOTELE PLATFORM

The architecture of the ARISTOTELE platform is made up by three logical layers:

- Tools layer: a tool is a software entity that provides a set of coherent functionalities that are offered by an user interface to the members of the adopting organization.
- Services layer: a service provides the business logic for the functionalities of the platform and the implementation of the business processes; each service might be used by the higher-level tools or invoked by another service [6], [8].
- Data layer: this layer includes all the information accessed and managed by the platform, either in structured, semi-structured, or unstructured form.

Basically, users interact with the platform via the higher-level tools, each designed for a specific purpose and developed accordingly. The functionalities the users are offered rely upon the lower-level business logic that is implemented by the platform’s services, including a specific service devoted to the management of data for the whole system.

Further details regarding these layers and their components are described below, starting from the data layer and moving upwards along the platform’s hierarchy.

A. Data layer

As we briefly stated earlier, this layer is made up of all the information managed by the platform. We can identify different kinds of data managed by ARISTOTELE, lying at different abstraction levels: content data and semantic data (placed at the base level), and business data (at a higher abstraction level with respect to the former two). Let us elaborate on those.

1) Content data: Content data consist of documents or any other kind of digital information, accessible by the members of the adopting organization: they can be either structured or unstructured, can reside in heterogeneous data sources and can be managed by different applications.

Within the context of the ARISTOTELE platform, content data are mainly coming from and stored in the respective databases of the two main reference frameworks adopted, i.e. SharePoint Server 2010 and Intelligent Web Teacher
(IWT) — more on them in Section IV-B. As a matter of fact, we consider data stored in SharePoint Server 2012 and in IWT as the primary information space for ARISTOTELE; however, additional data may also come from external data sources, and thus might need to be integrated with the primary information space.

2) Semantic data: Semantic data are used to represent the ARISTOTELE models and their instances. We will not delve much deeper into the ARISTOTELE models, since a thorough description of them is outside the scope of the present paper. Here, suffice it to say that they can define the context of interest for an organization’s project via a set of concepts and relationships between them, specified by means of standard dictionaries and ontologies. As we stated earlier in Section III, standard formalisms have been adopted for the representation of the semantics of information: Resource Description Framework (RDF) and Web Ontology Language (OWL). By means of these formalisms the information space takes the shape of a graph of semantically interconnected nodes.

The main idea behind these technologies is to define a set of concepts and relationships among them that represent the shared knowledge base for a domain of interest. Information resources are then annotated with these concepts by defining instances that provide a definition for what each resource is and how is connected with the other resources.

Therefore, in ARISTOTELE instances of semantic data are used to annotate content data, by relying upon the models previously created for representing the organization’s relevant concepts, as well as other concepts that have been introduced for the purpose of the specific project. This mechanisms eventually brings about the semantic network of information mentioned earlier in Section III, which in turn brings the Linked Data paradigm to fruition within the information space of the adopting organization.

Incidentally, semantic data get stored within the platform into an open-source Semantic Repository, Sesame[15], well-known amongst the Semantic Web community for its performance and scalability features.

3) Business data: Information managed by the tools are built by merging the different kinds of data mentioned in the previous paragraphs into encapsulating classes: we call them Business Entities, and they basically build up the business data of the platform. Business Entities have the role of decoupling the communication and management of data from the nature of the actual sources where those data come from, in terms of an object-oriented representation: for the purposes of the ARISTOTELE platform, they are expressive enough to correctly represent both the semantic and content data of the system.

B. Services

ARISTOTELE’s architecture has been designed according to the principles of Service Oriented Architecture (SOA).

The core idea of SOA is to design software applications as a set of interoperable services where each service is purposefully responsible of providing a specific, well-defined business functionality that can be easily reused; at the same time, services can be combined with each other to offer more complex functionalities.

By following these guidelines, in ARISTOTELE services implement the business logic and data access, and can be distinguished into Base Services and Core Services. Base Services encompass all the basic services from the reference frameworks and systems adopted by the platform that are relevant to it as a whole (e.g. User Authentication, User Profile Services, Business Connectivity Services etc.). Core Services, instead, are domain-specific and implement the actual business functions of the platform, and are instantiated to serve the upper-level Tools.

1) Base Services: ARISTOTELE’s Base Services draw upon the following reference systems:

- Intelligent Web Teacher (IWT) [4], [11]: IWT is an innovative, extensible and open platform for Learning and Knowledge Management conceived for delivering user-tailored and scalable e-Learning solutions.

Examples of Base Services from IWT are:
- Authentication: providing access to a IWT platform;
- Content Access: allowing content stored in the IWT database to be accessed by remote clients.

- SharePoint Server 2010: SharePoint Server 2010 is Microsoft’s business collaboration platform for the enterprise and the Web. It is an extensible platform for creating solutions, particularly suitable for the production of collaborative enterprise applications. Incidentally, the SharePoint user interface is inspired by the widespread Microsoft Office interface, in order to be more easily accepted by the users.

Examples of Base Services from SharePoint Server 2010 are:
- Authentication: providing access to a SharePoint site by using form-based authentication;
- User Profile: allowing remote clients to manage user profiles, colleagues and memberships;
- Access to SharePoint sites and lists: providing methods for the management of sites and lists;
- Search: allowing remote queries against the SharePoint server;
- Social Data: providing methods to create and manipulate social data;
- Microblogging: defining blogs for discussing various topics;
- Activity streams: monitoring and tracing the activity of the users;
- Security: managing security groups and authorizations remotely.
2) Core Services: ARISTOTELE’s Core Services are the following:

- **Knowledge Building Shared Services (KMSS):** The purpose of the KMSS services is to perform all the knowledge building capabilities of the platform, like text analysis, conceptualization, ontology building and ontology merging. Besides, KMSS implement the services related to specific business functionalities for the tools (i.e. expert finding, tag suggesting, competence gap computation), which leverage on the knowledge building capabilities themselves [2]. Specifically, these services present the following features:
  - Parsing & Natural Language Processing, including all the activities (services) for analyzing the content of resources (like Language Detection, Stop Word Removal, Stemming, Lemmatization, POS tagging etc.);
  - Data Mining, including a set of functionalities to extract conceptualize parsed text, group/cluster contents and establishing document hierarchy;
  - Similarity evaluation, a set of sub-services for evaluating different kind of similarities (syntactic, structural and similarity).
  - Ontology Labeling.

- **Linked Data Layer (LDL):** The LDL service enables data access for tools and other services, by allowing the aggregation of information distributed over heterogeneous sources and offering a set of storage-independent classes. By means of a Web Service, the LDL exposes a set of methods for managing and querying data that can reside on different repositories.

- **Recommender System (RS):** The RS service provides information filtering in order to recommend information items that are likely to be of interest to the worker. This is done by providing custom-tailored relevant information according to the characteristics of the users.

- **Social Networking (SN):** The SN service focuses on building and supporting social relationships among people following the principles of Enterprise 2.0 [5].

C. Tools

Tools build up the presentation layer: a tool is a software entity conceived and designed to provide a set of conceptually-related functionalities in order to solve some kind of real-world problem for a user. These tools access functionalities offered by services and provide the interface to the final users, and are developed by exploiting the SharePoint Server 2010 framework.

Specifically, the ARISTOTELE platform provides the following tools:

1) **Knowledge Management (KM):** This is a set of tools exploiting methodologies and techniques which support organizational knowledge building and maintenance (merging, matching, versioning) in a semi-automatic way, via the definition of knowledge extraction techniques as well. These tools are strongly based on the KMSS earlier described, and focus on activities such as:

- Tag Suggestion;
- Expert Finding;
- User-Generated Content Competence Analysis;
- Curriculum Vitae Analysis;
- Competence Gap computation;
- Team formation.

2) **Human Resource Management (HRM) Tools:** These set of tools provide functionalities for HRM activities, in order to identify the best set of human resources to be assigned to a project’s group/team, community of experts etc., as well as to define the optimum set of learning objectives to be assigned to each worker according to his/her own expertise. These tools take advantage of methodologies and techniques supporting competence gap analysis, team and group formation, competence development for internal resources and recruitment. More specifically, this set of tools include the following capabilities:

- a Decision Support System (DSS) aiming at supporting HR managers in making decisions related to the internal development of resources and recruitment activities [7];
- a team formation method supporting the identification of group of workers capable of executing a determined task or deemed suitable for a project;
- a competence gap analyzer providing information on competence gaps both at the individual and the organizational level. For some functionalities, the HRM Tools rely on the KMSS (for knowledge extraction, alignment and similarity computation) and the RS service.

3) **Collaborative Networking (CN) Tool:** This tool exploits methodologies and techniques to combine adaptive learning strategies with emergent, “non-adaptive” competence change, based on serendipitous exploitation of other people’s knowledge made available in social networks. Actually, this tool has been designed so that it might support the innovation process within the adopting organization.

4) **Learning Experience Generation (LEG) Tool:** This tool exploits methodologies and techniques to monitor a worker’s behavior, to provide workers with suitable contents and didactic approaches, and to generate, adapt and manage personalized learning experiences custom-tailored to the organizational objectives. It is thus capable of:

- generating and executing a Personalized Learning Experience custom-tailored to specific competences to be developed by a worker [1];
- handling social and collaborative learning activities;
- collecting, formalizing, sharing and reusing knowledge emerged during the execution of learning experiences;
- involving experts in the learning experiences;
• supporting authoring of learning content by using the knowledge extracted by previous learning experiences.

5) **Personal and reliable Work and Learning Environment (PWLE):** This tool is available online and accessible by the organizational user base, with the purpose of assisting a worker in his/her daily working activities, by acting as a methodological and technological integrator for most of the other tools (HRM, CN, LEG). PWLE helps knowledge workers achieve their objectives, supports learning and knowledge activities, connects workers with each other, and contributes to the organizational knowledge.

V. Conclusion

In this paper we have presented the semantic-based software architecture of the ARISTOTELE platform, a collaborative system for enterprise management. ARISTOTELE’s flagship feature lies in enhancing the learning and training progressions of workers within their respective organizations, thanks to the use of semantic models and tools to create, handle and store knowledge and help the workers share it with one another. We have motivated its compliance with a well-known framework for Enterprise architectures, detailed its building layers and elements (data, services and tools) and shown how its semantic approach permeates the platform as a whole.

VI. Acknowledgments

The research reported in this paper is partially supported by the European Committee under the Collaborative Project ARISTOTELE “Personalized Learning & Collaborative Working Environments Fostering Social Creativity and Innovations Inside the Organization” under Grant Agreement n. 257886.

REFERENCES


[4] Capuano, N., Miranda, S., Orciuoli, F.: IWT: A Semantic Web-based Educational System. In IV Workshop of the AIM@IA Working Group on Artificial Intelligence & e-Learning, held in conjunction with the XI Int. Conf. of the Italian Association for Artificial Intelligence (AIM@IA), 2009, pages 11-16.


