Proposal of a Digital Ecosystem Based on Big Data and Artificial Intelligence to Support Educational and Vocational Guidance

Essaid EL HAJI
LIST Laboratory, Faculty of Sciences and Technologies, Tangier, Morocco
Email: hessaid@hotmail.fr

Abdellah Azmani
Faculty of Sciences and Technologies, Department of Computer Science, Tangier, Morocco

Received: 09 March 2020; Accepted: 08 May 2020; Published: 08 August 2020

Abstract: The work in this article focuses on the modelling of an intelligent digital ecosystem for educational and career guidance for students and young people seeking their first job or retraining. To do so, the multi-expert system paradigm was used to aggregate the different expertises required for a good guidance, the multi-agent system principle was used to have a modular and easily scalable ecosystem. Indeed, the agents of the system communicate with each other using the FIPA-ACL language, in a collaborative vision, throughout the orientation assistance process to perform tasks such as proposing business sectors, occupations, training, and training paths. The ontologies of the Semantic Web have been used to have a complete semantic description of the shared information and to promote communication between the different software agents of the ecosystem. Big Data principles have also been deployed to manage and exploit structured and unstructured data from different data sources related to the guidance ecosystem.

The ecosystem modeled in this way has several innovative and powerful technological and scientific aspects. Thus, in terms of design and modelling, the proposed ecosystem considers all the actors and factors involved in the guidance process, including labor market trends. In technological/scientific terms, it is based on methods that allow it to be modular and scalable.

Index Terms: Digital Ecosystem, guidance, Artificial Intelligence, Big data, SMA, Expert System, Ontology.

1. Introduction

Guidance is a dynamic and pedagogical approach situated at the confluence of individual aspirations and constraints, the requirements of the labour market, organization and institutional rules. Its governance is complex and requires the intervention of several actors. They need to work together in pedagogical relationships and exchange information to help young people make the right choices. The parents of a young student have all the socioeconomic and personal information about their child, teachers can assess his or her educational abilities and skills, and the guidance counsellor has all the information on the pathways and conditions of transitions to their gateways. The combination of the pre-mentioned information and its proper exploitation can lead to a good guidance. However, good guidance also requires consideration of the labour market, the country’s development strategy and major projects and the control of the socioeconomic environment. The combination of the various players mentioned above, their data, knowledge and roles, forms an ecosystem of the value chain of orientation. Such an ecosystem can be modelled by its digital counterpart to foster and facilitate collaborative work between the different stakeholders involved in the orientation or reorientation of individuals to succeed in their future project or their socioeconomic reintegration.

It is in this perspective that this work, in this paper, aims to design and model an intelligent decision-support IT solution to assist young people in their orientation process. To do so, scientific and technological methods, essentially derived from artificial intelligence, have been aggregated in order to implement a decisional, efficient and robust solution constituting a digital ecosystem associating all the institutional actors and all the protagonists directly or indirectly concerned by the educational and vocational guidance.

For artificial intelligence, we have relied on an array of innovative and emerging technologies, namely: the multi-expert system paradigm to aggregate the different knowledge related to the orientation, the multi-agent system paradigm for a modular, distributed, changing and intelligent implementation, and the semantic Web ontologies. The latter make it possible to define the various concepts of orientation as well as the semantic links that tie them. The
Proposal of a Digital Ecosystem Based on Big Data and Artificial Intelligence to Support Educational and Vocational Guidance

The implementation of these technological methods involves not only the use of agent-oriented programming and knowledge-based programming, but also the use of several robust platforms and tools such as JADE, Protege, JENA, SPARCL and JESS.

The digital ecosystem, designed as such, is an online platform linked to the information system associated with the ecosystem of guidance whose actors are: young people, parents, teachers, counselors, professionals, as well as other legal persons (companies, public, semi-public and private institutions, actors of civil society) to interact and exchange information in order to put a work together with a solid knowledge base.

2. Global Architecture

The proposed ecosystem aims to make a concrete and relevant contribution to the question of orientation in Morocco. Its design emanates from our reflections and our in-depth studies on the Moroccan guidance system, but also guidance on other systems in the world, especially in developed countries. As Figure 1 shows, the ecosystem involves several actors. Thus, it makes it possible to exchange data and knowledge with the various third parties having direct or indirect links with the question of orientation, namely, national education, higher education, vocational training and various ministries and public institutions as well as socioeconomic actors and those of civil society without forgetting, of course, the professionals and citizens directly concerned.

Indeed, this ecosystem represents a space for exchange and collaboration between several actors (pupils, students, parents, teachers, job seekers, guidance counsellors and professionals) and whose objective is to help young people, job seekers or those seeking retraining to find their way around more easily.

The proposed digital ecosystem is designed to foster guidance governance based on an architecture that links all the information systems of the different institutional players involved in guidance. It allows the various protagonists to access it for various uses (to inform themselves, to file a request, to use the e-window, to read practical advice, to make tests, to create a route, to develop a project for the future, to find out about a trade, looking for a player in training, building a skills assessment, discussing and exchanging information with other individuals, consulting statistical data, seeing predictions and many other types of use). Such a digital ecosystem is based on the principle of "open data", thus promoting availability, reuse and distribution, and universal participation.
The ecosystem provides decision support for two categories of people to be referred through two types of interventions: those who already have an orientation project and those who do not. For the first category, assistance consists in assessing the degree of correspondence between the choices expressed and the profile of the young person, after which the system decides whether the young person should continue with his project, and what are the possible actions (possible training, skills to be acquired, etc.). For the second category, the system offers an orientation project in four stages: choice of activity sector, choice of profession(s), choice of training(s) and choice of training path. This dynamic requires the intervention of 4 four experts according to the principle shown in Figure 2.

![Fig.2. Dynamics of generating a guidance project](image)

In this dynamic, the ecosystem offers a business capital based on their professional tendencies (Professional interests, professional sub-interests and preferred activity sectors). Then, this business capital is subjected to three evaluations before being validated:

- Personality trait assessment to test the correspondence between the young person's personality traits and those required by the job. If these traits coincide or are similar, the trade is maintained, otherwise it is expelled from the list of choices;
- Evaluation from a pedagogical point of view to test whether the young person's pedagogical possibilities (skills and training) allow him/her to exercise the chosen profession(s) or not, with possibly the tasks to be fulfilled (training, diplomas, ...) to approach the requirements of the profession(s) previously selected;
- Assessment from an economic point of view (labour market) in order to avoid directing young people towards jobs that are saturated, being saturated or disappearing.

For the implementation of such a dynamic, we have called upon the decision-making approaches we present in the following.

A. Approach based on the multi-expert system paradigm

In guidance field, the sources of knowledge are diverse and combine several areas of expertise (pedagogy, psychology, sociology, economics, etc.), which must absolutely be taken into account in order to implement an efficient solution. This knowledge can be modeled in the form of decision rules, hence the importance of using the expert system's paradigm.

According to J.C Pomerol, an expert system is a concept of artificial intelligence, designed to simulate the know-how of a specialist, in a precise and well-defined field, through the use of a certain amount of knowledge explicitly
provided by experts in a specific field. [1].

Generally speaking, an expert system can be defined as “a set of programs with which one can represent knowledge of one or more human experts in a given field and who is capable of operating a certain number of reasonings in order to make diagnoses and provide decision support in solving problems in said area”.

Sometimes the problem we are trying to solve is complex and requires the intervention of several areas of expertise from different fields. Hence, the need to create a system bringing together several knowledge bases, each in its specific field, and to enrich it with new rules related to inter-domain interactions. To solve this kind of problem, we use the principle of multi-expert system which allows aggregating several areas of expertise, and therefore calls on several experts to build the knowledge base.

i. Guidance factors and key experts

Educational and vocational guidance is a complex and multidisciplinary process requiring the combination and aggregation of several sources of information, knowledge and expertise, but also the consideration of an arsenal of important factors. Numerous studies have highlighted demographic and personal factors affecting the choice of vocational and educational guidance. Personal factors include general ability [2] culture or cultural status [3], values and principles [4], self-esteem [5], self-efficacy, interests [6], personality [7] are determining factors. Demographic factors include parental work [8], parental education [9], ethnicity [3], socio-economic status, gender and age of the student [2].

The multi-expert system that we propose attempts to address the various factors mentioned above by basing it on four areas of expertise: expertise in pedagogy (pedagogues, educators, teachers ...), expertise in profiling (guidance counsellor, psychologist, sociologist, coach ...) and expertise in economics (economists and professionals). In addition to these three areas of expertise, there is also expertise in follow-up and coaching (Tutor). These experts share a collaborative knowledge base fed and updated through several sources of knowledge. The organization of our multi-expert system is shown in Figure 2.

ii. Experts collaboration

The four domain experts work in collaborative mode and each of them has its own set of rules. The synergy, produced by the combination of the different expertises, makes it possible to better help a person to orient (especially young people) to make the right choice for their future project combining a profession, the necessary training and an apprenticeship path. It is the one best suited to his or her profile, skills and character, taking into account the current and future socioeconomic context of the region where he or she is (or wishes to be) located, as well as the labour market of the region and the national and regional employability strategies.

Fig.3. Experts of the system

The profiling expert is primarily responsible for identifying the interests and professional tendencies of the person to be oriented, based on the Holland RIASEC model, as well as the personality traits of the person to be oriented, with reference to the Big Five model. Following this identification, the profiling expert proposes, secondly, to the individual the most suitable business capital.

The pedagogical expert is in charge of evaluating and examining the professions proposed by the profiling expert in terms of educational curricula and then presenting his recommendations resulting in an acceptance (direct or conditional) of the profession or its rejection. In the case of conditional approval, the pedagogical expert must specify the conditions to be taken into consideration by the person to be oriented. These conditions may consist, for example, in

Copyright © 2020 MECS
the acquisition of a new skill and/or the improvement of other skills. This expert is also in charge of generating training courses and training paths with reference to the chosen profession(s) and pedagogical information concerning the individual to be oriented.

The economic expert evaluates the professions in terms of their situation in the labor market, particularly in terms of employability, wages, comfort, security and promotions. This assessment results in a detailed report on the job(s) concerned. This expert also suggests to the individual to direct suggestions (skills required or to be worked on, related professions ...).

The Tutor expert is responsible for monitoring the individual’s project. To do this, it constantly “listens” to the system in search of updates, whether for the individual (new skills or new abilities, for example) or in the collaborative knowledge base (change of rule for a trade, a training course, etc.) and to advise the person to be referred afterwards.

The proposed ecosystem is organized, therefore, around a multi-expert digital system for educational and vocational guidance and counselling (SMENO: Système Multi-Expert Numérique d’Orientation), which brings together the various areas of expertise already available. SMENO includes all the necessary elements for the retrieval of expertise, its processing, communication and explanation of the results.

---

**B. Multi-agent’s system approach**

In recent years, centralized IT systems were beginning to reach their limits and were gradually withdrawing in favor of new systems based on the multi-agent paradigm, more suited to a modular and distributed architecture [10]. Indeed, computer systems are becoming more and more complex and use distributed and heterogeneous resources, hence the need for decentralization and cooperation between different software agents to solve a problem.

We are, therefore, intuitively led to look for new ways to give more autonomy and initiative to the different software entities. The multi-agent paradigm proposes a framework for responding to this challenge and according to which complex systems can be broken down into several units that communicate to solve a problem.
i. Multi-layer architecture

The multi-layer representation allows a high degree of abstraction of the solution and allows a large margin of freedom in the choice of programming language, databases and technologies to be used. Such a presentation also makes it possible to clearly delimit and situate the problems. The model we have chosen opts for three main layers: Human Machine Interface Layer, Adaptation Layer and Semantic Layer. These three layers are schematized in figure 5.

- **IHM Layer:** It represents the communication channel between the users and the different elements of the solution. The GUI allows users to authenticate themselves and perform the various tasks entrusted to them. It is managed by a set of software agents.
- **Adaptation Layer:** It is the main layer of the solution in the sense that it carries out the treatments and draws conclusions. This layer is based on a set of artificial intelligence methods such as software agents (associated with SMA also called virtual agents), the FAHP method, K-nearest neighbors, neural networks, etc. The tasks performed by this layer are the responsibility of a group of intelligent software agents.
- **Semantic layer:** This layer is made up of ontologies allowing semantic modeling of different knowledge. This layer is in turn managed by a set of software agents.

![Layered representation of the proposed architecture.](image)

Fig.5. Layered representation of the proposed architecture.

ii. The agents of the system

The identification of agents in a complex and open system is a delicate task and requires a deep reflection on the number of agents, the tasks of each agent and the possible interactions between them. The relevance of the choice of agents therefore requires a perfect understanding of the system to be designed. For our solution and after an in-depth study, we opted to have 20 agents, each of whose role is briefly presented below. Each agent operates at the level of one of the layers of the architecture. The different agents communicate with each other via the FIPA-ACL agent communication language, which is the most widely used and studied language and it incorporates many aspects of KQML [11,12]. Figure 6 shows an illustration of the proposed MAS.

- Agent Interface Candidate (AIC): the main role of this agent is to receive the different requests from a candidate, to transmit them to the Supervising Agent Manager and to communicate the answers to the candidate;
- Agent Interface Expert (AIE): This agent manages queries generated by domain experts, such as updating decision rules;
- Agent Interface Professional (AIP): this agent allows the system to communicate with professionals in order to give their opinions (evaluation) on trades and training;
- Agent Manager Supervisor (AMS): this agent coordinates the work of the various agents and ensures that the system works properly;
- Agent Professional Interests (APr): This agent identifies the professional interests of the candidates via the Holland R-I-A-S-E-C test;
• Agent Personality Traits (ATP): this agent identifies the dominant personality traits of the candidates. To do so, he uses the BIG five test as a basis;
• Agents Evaluator Pedagogic (AEP): the mission of this officer is to evaluate the disciplinary skills of the candidates. This assessment is important for training choices;
• Agent Evaluator Marché du travail (AEMT): this officer is an expert on the labor market and has a good knowledge of the present and future situation of the labor market. He/she is able to rate a job in the market in terms of employability, salary, social security, promotion and comfort;
• Agent Calculation Similarity (ACS): this agent allows a candidate to compare himself/herself to a typical profile by calculating the similarity between the acquired profile and the required profile. It can place a candidate in one of three states: sufficient similarity, partially insufficient similarity and completely insufficient similarity;
• Agent multi-criteria analysis (AAMC): this officer uses MCDM methods (in particular FAHP) to propose the best possible alternatives to an orientation candidate, including business sectors and training institutions;
• Agents Trades Generator Métiers (AGM): The role of this agent is to determine a business capital that best corresponds to a candidate's profile. To do this, he communicates with the AIP, ATP and AEP agents via the AMS agent;
• Agent Training Generator (AGF): It offers a range of training courses in line with the business capital of the company. ;
• Agent training paths Generator (AGP): it identifies the possible training paths for a candidate in line with the Skills capital and the Training capital;
• Agent Tutor (AT): It listens to the system for possible updates. It sends out notifications to inform an applicant in the case of changes that may occur as a result of changes in the environment (updating an applicant's data such as having new skills, for example, a change in the status of an occupation in the labor market, etc.).
• Agent Knowledge Management (AGC): this agent allows you to edit the knowledge in the collaborative knowledge base;
• Agents Evaluation and Filtering (AEF): this agent is solicited by other agents, in particular the ATP, AIP and AEP agent, when determining an orientation project (Job, training and career path). He or she searches the knowledge base for the best matches between a candidate's profile and the various profiles required.

Fig.6. Representation of System Agents.
C. Approach based on the use of semantic web ontologies

In order to better identify the different concepts involved in the guidance process and the relationships between them, we have opted for the ontological approach. Moreover, the ontological approach allows us to build a concept that is easy to maintain and to make evolve afterwards.

The literature covers a variety of definitions of the term "Ontology" that vary from one disciplinary field to another. In the world of knowledge engineering, according to Gruber, an ontology is a formal and explicit specification of a shared conceptualization, i.e. an ontology is a description of the concepts and relationships that may exist for an agent or a community of agents [13,14]. Typically, ontology contains a hierarchical description of the important concepts in a domain and describes the properties of each concept through a value-attribute mechanism [15].

i. Related Works

Before starting to think about our own ontology, we did some research to see if there is an existing ontology that converges with our work in order to benefit from it by adapting it to our context. Unfortunately, we found only examples without any real convergence with our topic. As an illustration, we cite the work [16] in which the authors propose a competence-based ontology for the annotation of CVs/Job Offers and the work of a PhD thesis [17] in which the author proposed an ontology for the generation of personalized learning paths in the framework of an e-Learning project.

ii. Overview of developed ontologies

Not having found an existing ontology that could satisfy us, we developed our own, making our work more original and more open to other researchers. In order to set up an ontological model that is easy to evolve and maintain, and to better optimize it for potential queries, we have chosen to divide our ontology into six sub-ontologies. The figure 7 presents an overview of these ontologies and their interactions:

Fig.7. Overview of developed ontologies

- Ontology of trade (OntoTrade) which provides the vocabulary necessary to describe the elements of a trade. Trades are classified by industry sectors. This ontology is related to the Competency ontology (OntoCompetence) to certify the skills required by a trade;
- Diploma ontology (OntoDiploma) which provides the terminology necessary to describe the degrees that a trade may require or that a candidate may have. These are categorized according to the cycle of study. This ontology is linked to the Discipline sub-ontology (OntoDiscipline) to determine the discipline of a given degree and to the Competency sub-ontology (OntoCompetency) to attest to the competencies mobilized by a particular degree;
- Competency ontology (OntoCompetence) which describes the general and specific competencies that a professional activity may require or that an applicant may have. This ontology is linked to the sub-ontologies OntoCandidat, OntoDiploma and OntoTrade;
- Candidate ontology (OntoCandidat) that describes the profile that an applicant may have. It is linked to the OntoDiploma, OntoTrade and OntoCompetence sub-ontologies.
- Ontology of Discipline (OntoDiscipline) which describes the branch and discipline of a degree. It is related to the OntoDiploma and OntoCompetence sub-ontologies;
- Career Ontology (OntoPath) which describes the educational pathways a candidate may take. It is linked to the OntoDiploma sub-ontology.
D. Using Big data approach

The proposed ecosystem is designed with an ambitious perspective of using data from multiple sources; therefore, the integration of a layer dedicated to big data and data science has been intuitive and natural. This layer will allow us to receive and exploit the big data to extract solid and up-to-date knowledge about the subject of guidance. This knowledge focuses on the actual situation of the labor market and its trends in terms of skills and abilities demanded, profiles sought and those in saturation or even deterioration. The information collected also relates to the country's strategy and orientation in terms of hiring, development and major development projects. More concretely, Big data principles were used to automatically generate decision rules for the expert system's knowledge base through the use of machine learning algorithms applied to the collected data.

The main components of the big data layer related to the proposed ecosystem are as follows:

- Datawarehouse External: to store data from multiple sources of information from the stakeholders listed in Figure 1.
- Web of data and social media: the advent and development of the social web and associated technologies, including social networks, online recruitment and profiling platforms have opened up horizons towards a world of knowledge sharing and co-production and have opened the doors to communities of interaction, exchange and sharing of information, expertise and opinions. The exploitation of data from the web using artificial intelligence techniques and machine learning algorithms will allow us to make good conclusions about labor market trends at local, regional, national and international levels.

3. Technical Architecture

Technically, it is a multi-expert system supervised and organized around a multi-agent system, which is based on the n-tier model of a full web application. Figure 9 shows the components of the system and the interaction between them. This system builds its collaborative knowledge base using different ontologies. On their side, the different software agents have access to the ontologies implemented via the JENA framework and the SPARQL language.
The system uses the Protege editor to edit the ontologies, the framework JENA and the SPARQL query language to edit the OWL file; JADE allows the implementation of SMA and JESS which is an environment and rules engine written entirely in Java by Ernest Friedman-Hill at Sandia National Laboratories. In a multi-agent environment, JESS is used as a decision element for an agent to make decisions.

Concerning the Big data layer, HDFS (Hadoop Distributed File System) is dedicated to data storage; it is mainly used for distributing the data across the nodes, managing replication for redundancy and administrative tasks like adding, removing and recovery of data nodes. HBase is a distributed, column-oriented DBMS that provides real-time read and write access to data stored on the HDFS. Sqoop (SQL-to-Hadoop) is a tool to transfer data from a relational database to the HDFS of Hadoop and vice versa. Flume allows importing and transferring high-volume data. Pig and Hive are used to query the data.

4. Conclusion and Perspective

This work is part of a more general perspective that aims to make a scientific contribution to the question of educational and vocational guidance, particularly in Morocco. Its objective was to model a digital ecosystem for educational and vocational guidance, especially since employability is at half-mast and training and diplomas no longer meet market needs and do not really converge with the country's sectorial plans and development prospects.

In order to set up an efficient and evolving digital ecosystem, many methods and technologies, essentially derived from artificial intelligence, have been studied and developed, namely the multi-expert system paradigm, the principle of multi-agent systems and semantic Web ontologies. In addition, big data techniques allow processing massive data related to the information systems of different actors of the ecosystem, but also to exploit the potential of the data web and social networks to extract relevant information in real time and build a solid knowledge base for the multi-expert system that represents the core of the ecosystem.

Beyond what is being done, this work offers several prospects for research and development, particularly within the context of big data and data science, to make the proposed ecosystem a powerful tool for guidance and work search assistance.

References

Authors’ Profiles

**Essaid EL HAJI** (Morocco, Tangier, 1981) received his Ph.D in computer science and artificial intelligence at faculty of science and technology, Tangier-morocco. His current research interest includes artificial intelligence methods and technics in educational and career guidance. He is a Teacher of computer science at the Chaabane High School in Larache-Morocco. He has several papers in International Conferences and Journals.

**Abdellah AZMANI** (Morocco, Tangier) received his Ph.D in Industrial Computing at the University of Science and Technology of Lille in 1991. He worked as a professor at the Ecole Centrale de Lille and at the Institute of Computer and Industrial Engineering from Lens. He is a member of the Laboratory of Automatics and Informatics of Lille (LAIL). He is a professor at Faculty of Sciences and Technology of Tangier, Morocco. He has contributed to many scientific researches.