Final Project

The Performance of Knowledge Organizations and Modelling Human Action

Quantitative Models for Economic Analysis and Management

Sapienza University of Rome

Group C

Simone Boesso

Davide Cacciatore

Giulio Corsetti

Valentino Sacco

Braulio Villalobos

Executive Summary

This document aims to give a clear introduction and representation on the way in which modelling human action has become essential within the knowledge organization field to achieve high performances. Starting from the Systematic Review carried out over the just mentioned topic, we will expose how our ontological model supports this statement in a clear and precise way. At the end, we carry out an empirical analysis on the dataset *Digital Economy and Society Index (DESI)* by using the *Visual Analytics Environment (VAE)*, which will serve the purpose of understanding the interaction between knowledge organizations, human action modelling and several other related components extracted presented in the Systematic Review and the Ontology.

This document actually contains all the main ingredients needed to develop a model: theory, method and data. Looking at the basic definitions from the Oxford dictionary, we know that a model is a simplified description of a system or process, to assist calculations and predictions. We think that the model we developed trying to explain the connection between human actions and knowledge organization has all these important characteristics.

The first ingredient is the theory, defined as a system of ideas (general principles) intended to explain something. The course of Quantitative Models for Economic Analysis and Management gave us all the theory we needed to complete this analysis through different methods and point of views.

The second ingredient is the method, defined as a series of steps or acts taken to acquire knowledge or perform a task. It is a particular procedure for accomplishing or approaching something. The main method we used for this project is the Systematic Review that allowed us to collect a large amount of information on the topic and to learn what is the state-of-the-art of the knowledge organization and what are the possible improvements.

Finally, we have the data, defined as facts and statistics collected together for reference or analysis. At the end of the document we provide an empirical analysis on real data to gather some relevant information about the knowledge organizations and how they can affect, or they are affected by, human behaviour.

In particular, the chosen dataset looks at the European countries and contains many indicators which are useful to understand the behaviour of public and private organizations in the digital era. In our opinion, there is a need to understand the real impact of Big data on economic modelling and management. As we have read in the book *Management - Revised Edition (P. F. Drucker, J. A. Maciariello)*, the data plays an increasingly central role into the decision making process of a good manager. There are several articles and examples in the Systematic Review that talk about this argument.

In conclusion, we have been able to learn the topic of the knowledge organization from a “digital” point of view. We focused on the new technologies applied to the information retrieval processes. There are countries that have been able to improve themselves and others that need a few more steps. Anyway, the take home message is that the digital innovation in modelling human action and in the knowledge organization mechanism will be more and more crucial and indispensable.

We understood that using modelling in a knowledge organization context can improve the working performance both of a private and public firm and also the human performance.

Systematic Review

Abstract

This systematic review aims to highlight how modelling is a useful tool to get better performance in knowledge organization. According to *The Organization of Information* by Joudrey and Taylor, information organization: "examines the activities carried out and tools used by people who work in places that accumulate information resources for the use of humankind, both immediately and for posterity". The traditional approaches performed by librarians, archivists and specialists are increasingly challenged by computational algorithmic techniques that deal with Big Data. What we really want to investigate in this review is the contribution of a good knowledge management system and of a good knowledge organization to firms and humans' performances.

Introduction

Knowledge organizations' success depends on how they manage their knowledge, in fact only being able to improve this capacity they can innovate and survive in the market. In this systematic review we focus our attention on how modelling can be used to improve performance and, in particular, on how it's possible to model mechanisms where the human action is an input.

Methodology

To do our research we extracted specific keywords from the area of "knowledge organizations". In addition, we add "model" and also "performance" to restrict the domain and answer our question better.

We limited our research only to articles in "English" and we considered only the ones from 2010, in order to get new and updated articles. Thanks to these limitations it was easy to find several results among international papers, academic research, and transcriptions of international conferences.

To retrieve these papers, we used the SCOPUS portal. We implemented a simple PICO model to formulate in a structured way our question.

	Population	**I**ntervention	**C**omparison	**O**utcome
Key concepts	Knowledge organization	Non-parametric	All methods	Models and performances
Natural language terms	Knowledge Organi?ation*			Model* Performance*

More precisely, we used the following query into the advanced research:

(TITLE-ABS-KEY ("knowledge organi?ation*") AND TITLE-ABS-KEY (model*) AND TITLE-ABS-KEY (performance*)) AND PUBYEAR > 2009 AND (LIMIT-TO (LANGUAGE , "English"))

It is relevant to notice that we used the keyword *organi?ation* to check for both the words organization and organisation. In addition, we added * at the end of each word in the query, so we can also consider the plurals of the given words.

From this advanced research we obtained 65 papers. Then, we joined this group of papers with another group of 41 papers obtained with the Snowballing technique using Connected Papers from the article "*Designing a maturity model for analyzing information and knowledge management in the public sector*" by Aki Jääskeläinen. In our opinion, this article was really on target with our analysis, so we don't want to miss it and we thought that looking into its connected paper would be interesting too. We can see them into FIG. 1 in the Appendix.

We collected all our obtained documents with the software Mendeley Reference Manager, that allowed us to share our bibliography and to work together at the same time, and we removed 10 duplicate documents between the two groups.

We followed the PRISMA method to select the final relevant studies that we have included in the review, as reported in the PRISMA flow diagram [FIG. 2] in the Appendix.

At the end, we obtained 96 papers. We are a team of 5 members, so we searched 19 documents each.

After a quick review of the titles and of the abstracts of all these articles, we decided to keep 9 documents in total, as some of the other articles were not strictly related to our research.

Results

At the end of our research, we obtained a total of 96 documents, 55 from the query performed on the SCOPUS database and 41 connected papers to the article "*Designing a maturity model for analyzing information and knowledge management in the public sector*" by Aki Jääskeläinen.

In the "Appendix" section, at the end of the review, are shown the bibliographic analysis that we have made using the R package Bibliometrix for the file exported from SCOPUS and from the Connected Papers web tool.

In FIG. 3 there is the Co-Occurrence Network of the 65 articles found in SCOPUS. We can see that the most used keywords are knowledge management, knowledge organization and human. There are semantic clusters of words related to "knowledge", "human" and "learning" keywords.

By applying a clustering algorithm on the keywords, it is possible to highlight the different themes of a given domain. Each cluster can be represented on a particular plot known as Thematic Map (Cobo et al., 2011) [FIG. 4]. On the x axis we have the centrality which can be read as the importance of the theme in the entire research field, while on the y axis we find the density which can be read as a measure of the theme's development. The plot shows us that the motor themes are led by three clusters of keywords. The blue one is about human actions, the green one is about knowledge based systems and the red one is about knowledge organization and management. It means that these topics have a high relevance degree inside our research and, at the same time, they have a high degree of development.

If we divide our time span (2010-2022) into two equal time slices, (2010-2016 and 2017-2022) it is possible to study and plot the topic evolutions into a longitudinal Thematic Evolution Map [FIG. 5] as a Sankey Diagram. In the first slice, the central topics were knowledge management, artificial intelligence, human, knowledge based systems and software engineering. If we look at the evolution in the second slice, we can notice that knowledge management and artificial intelligence combine themselves into the knowledge organization topic. Moreover, the knowledge based systems evolved into ontologies. So, we can affirm that the artificial intelligence applied to knowledge management returns all the knowledge organization processes, while the ontologies became the more important and more used kind of knowledge based systems.

Organization Capital

For a knowledge organization the ability of innovating and its performance depend on the organization capital.

Organizational capital is the value to an enterprise which is derived from organization philosophy and systems which

leverage the organization's capability in delivering goods or services.

We start focusing on the link between Knowledge Management and performance.

What is Knowledge Management and how it is linked to organizational performance in companies?

In *"Knowledge Management, Knowledge Management System, and Organizational Performance: An Empirical Study"* the authors W. Han and Y. Wang analysed the relationship among these three factors. Knowledge is the determinant factor to keep the competitive advantage of enterprises and the purpose of this study is to examine how the usage of it can support management and organizational performance. But first, let's define what knowledge management is.

Knowledge management is a dynamic process consisting of many parts.

1. ***Knowledge creation***. Knowledge creation is the capability of developing valuable new thoughts and schemes.
2. ***Knowledge organization***. Knowledge organization involves several processes: storage, maintenance, retrieval and protection of knowledge.
3. ***Knowledge transfer***. Knowledge transfer is a specific process by which knowledge is transformed, transmitted and absorbed by users. It consists of three dimensions: acquisition, dissemination and absorption.
4. ***Knowledge application***. The competitiveness of an organization depends on how effectively it can apply and integrate the existing knowledge in the decision-making process.

The model proposed by the authors verified the relationship between knowledge management (KM), knowledge management capability (KMC) and organizational performance. KMC can improve the learning ability at individual, group and organizational level. The higher learning ability will support the organizations to absorb new knowledge and integrate new and existing knowledge. KM practices can enhance new products and services development processes, reduce the time-to-market and increase the probability of innovation success. In addition, KM can improve the adaptability to market environment change and ability to make quick responses to new business opportunities.

This paper focuses on verifying different hypotheses on the relationship between KM and organizational performance. They conducted a study in which 176 organizations with KM practices were involved.

They found that knowledge application impacts directly the organizational performance instead of using the KM capability as mediator. Then, the use of IT will lead to more effective knowledge creation, organization, transfer and application but it will not directly affect the KM capability and they will not improve the organizational performance.

The limitations of this research are the focus only on technical factors. There are also some non-technical factors like culture and organization structure that are not considered.

Applying Knowledge Management to Human Performance

Apart from allowing enterprises to coordinate and develop efficient routines to enhance working performance, KM can and should be applied to human performance: when we think of an enterprise often we focus only on the higher end of the

production chain, but John Dumay's paper "*Using critical KM to address wicked problems*" shows how KM should be used to address wicked problems amongst those who contribute to provide the raw product to the higher level companies that apply KM to enhance their productivity.

This issue has not been covered often, but articles that treat it mostly recommend that researchers should change their approach to both research and practice: concentrating too much on organizational performance can distract from the whole, and other than to asking us "What is KM made up of?" we should ask "What should KM do?".

The Finnish Model

An example of this change of approach can be found in the paper "*Knowledge management and hybridity of institutional logics in the public sector*" from H. Laihonen and P. Kokko, which focuses on the Finnish journey of shifting public administration to a hybrid approach, making data public and involving organizations in administration.

At the starting time of the study, Finland's goal was to enhance productivity of the public sector starting with the Isaacus Project. This programme raised knowledge management to a national agenda and in 2015 this was also extended to administrative branches.

The Data Lake

Apart from improving health care and social provision, the purpose of the Isaacus project was to make knowledge accessible to most actors, enabling scientific research, statistics and effectively shifting the focus from organizational information to platform-based thinking. This was achieved by introducing a new technology: the "data lake".

As cited by J. Dumay, according to Callon, M. (p. 206, 1986) the common interest is an "obligatory passage point" that bonds all companies from the same fields, because "they cannot attain what they want by themselves" but could, by forming coalitions, tackle problems that are way too big for one to solve.

This is exactly what Finland has done by introducing Data Lakes: many actors managed to learn and provide knowledge on shared topics; all the interviewed parts in the study were seen to have learned a lot and gained in performance.

Observations on the Finnish Model

This project highlights how a new information structure could challenge and reform the whole institutional order and shows how thinking about KM as a "problem-driven discipline", breaking the individual enterprise boundary, could improve organizational performance by forming collaborative coalitions capable of approaching and solving major problems.
Effectively, this change of approach could not only enhance working performance, but *human performance.*

An Organizational Capital Decision Model for Knowledge-Intensive Organizations

We highlighted the link between knowledge management and performance, however, to have effective management we need to deploy the right organizational capital.

Usually, when we think about capital, we think about the capital used to get the physical needed resources to create a product/service, but especially for the knowledge organizations, their success depends heavily on their capacity to innovate.

Therefore, we can't decide approximately how many resources we should give to this fundamental area in a firm. We need a model to analyse and make this decision. A model used to answer the following question: How to allocate the budget between physical capital and organizational capital?

Organization capital, human capital
To define this model, we need first to define the roles and what aims this capital should go for, but in a real context we don't always have a clear boundary, and we can't find little pieces of information about what is and what it includes from firms' reports and accounting books.

Therefore, the first step is to define these boundaries, and using the definitions proposed in "*An Organizational Capital Decision Model for Knowledge-Intensive Organizations*" we define human capital as the contribution to work output affiliated with the amount of knowledge in the firm (but also new sources like DB, patents) and organizational capital as the contributions to how human capital creates innovation.

The model: the authors suggest a model for asserting optimal allocation to organization capital maximizing expected income, to analyse potential benefits and expected income.

They implemented this model using a Monte Carlo simulation with also real data from the top global software firms. They used these companies because software companies heavily operate on the main quantitative elements defining organizational capital, such as training, the knowledge integration mechanisms and organizational capabilities. Additionally, various aspects of the software industry are ubiquitous in several other industries, such as online retailers, and transportation.

Results analysis: They proceed to analyse these companies dividing them into 2 categories: "*High delta / Low delta*" depending on their optimal allocation of total budget to organizational capital. They compared the expected income using this model and its optimal allocation with the number reported in the financial filings. They discovered the firms with the highest optimal delta would have almost doubled their income if they used the suggested model.

The importance of information retrieving

The fundamental ability of a knowledge organization is retrieving knowledge efficiently from the resources, and therefore it's one of the mechanisms where a firm should invest.

M. A. A. Leite and I. L. M. Ricarte studied how knowledge organizations can improve their performance in retrieving information based on users' queries. In "*Relating ontologies with a fuzzy information model*" they present a novel way to represent and organize knowledge, from distinct domains, using multiple ontologies that can be related.

A basic information retrieval system stores and indexes documents in such a way that, when users express their information needs in a query, the system retrieves the related documents assigning a relevance score to each one. Usually, documents are retrieved when they contain terms or keywords specified in the user's query. However, many other documents may contain the desired semantic information, even though they do not contain the user-specified keywords, in which case the traditional approach leaves out relevant documents. One way to deal with this limitation is to consider not only the lexical information explicit in documents but also its semantics.

Each domain can be represented as a conceptual structure, such as a lightweight ontology. They include concepts, concept taxonomies, relationships between concepts and properties that describe concepts.
Relationships between concepts can be represented by fuzzy relations, with crisp relationships being a particular case.

The model evaluation uses a 129-document collection sample of the agrometeorology domain in Brazil; a 83 query set; a lightweight ontology of the geographical Brazilian territory and a lightweight ontology

of the climate distribution over the Brazilian territory. Both ontologies are manually constructed.

The query set is composed of queries containing a single concept from each ontology, as well as queries containing combinations of two concepts from both ontologies connected by AND or OR Boolean operators. Each document is examined and reassigned to the ontology concepts by a domain expert.

The knowledge representation and the method proposed for query expansion were tested using the Apache Lucene Engine. After that, the performance was compared with the multi-relationship fuzzy concept network information retrieval model. It has been shown that the proposed model performs better than the others. For fuzzy ontologies, the precision for low recall values is above 95% and remains above 50% for higher recall values.

The innovation of this approach is that the ontologies can deal with knowledge of distinct domains and they can be related to each other by casual, spatial or similarity relationships.

They demonstrate that human actions are needed to have good models for knowledge organization and they proposed a new information retrieval system that performs better than the others.

Sense induction in folksonomies

A real complex aspect of information retrieval is the presence of folksonomies.

Folksonomies often known as tagging system, widely used in many websites, use a very simple knowledge organization system, this allows to create quick extensive knowledge annotations, but it incurs in some barriers for the automatic use of such knowledge organization systems by computers and new techniques must be developed to extract the semantic of the tags used.

The authors Pierre Andrews, Juan Pane and Ilya Zaihrayeu (DISI - University of Trento, Italy) had explored new ways to deal with this complex aspect.

Introduction
Folksonomies are described as uncontrolled knowledge organization systems where users can use free-text tags to annotate resources. In fact, they create a network of user-tag-resource triplets that encodes the knowledge of users. This structure however must deal with language ambiguity issues and synonymy which represent a great problem for computer algorithms.

Word Sense Disambiguation unfortunately are not completely adapted to folksonomies: they use an existing vocabulary to link terms (in our case tags) to concepts, thus discovering the semantics of the tags used.

Sense induction
The method used to extract the semantics from folksonomies is what is called *tag clustering* and its principle is based on machine learning clustering algorithms. This clustering is based on the principle that similar tags will have the same meaning and can thus be attached to the same “concept” in the created vocabulary. However, to compute the similarity between tags to run the clustering algorithms that will attach similar tags together remains the main issue. For instance, the authors claim that all the methods available currently use a mix of measures based on the collocation of tags on resources and their use by users. If two tags are often used by the same user on different resources or by different users on the same resource, then they can be considered similar. Since tags are often ambiguous in folksonomies and can bear more than one meaning this is the main weak point. In the article to obviate this issue it is added an extra step to the clustering to first identify the diverse senses of polysemous tags and in the following

clustering steps, we do not consider tags directly, but the unique senses that they can take.

The algorithm
In the article the well-known KNN and K-means algorithms are avoided since they don't match the purpose: the step in which clustering is applied to detect the different senses in which one tag can be used. In this case, it is not possible to find an overall optimal value for the number of clusters to look for as each term might have a different number of senses.

For instance, the DBScan algorithm is used to do a density-based clustering. This approach to clustering has various advantages for our application:

- It does not require as input the number of clusters to be found.
- The DBScan algorithm can decide that some of the items to be clustered are noisy and should not be considered.
- The DBScan algorithm can detect clusters that have more complex "shapes".

The authors in the end present their algorithm which is divided in 3 steps:

1. For each tag, we cluster the user-resource bipartite graph that is attached to this tag.
2. Then apply the same principle as the one discussed in the state of the art on the user-resource-sense tripartite graph to cluster similar senses together.
3. Identify new concepts for each of the clusters. This process is equivalent to finding the relation of the new concept in the structured vocabulary.

The Authors then affirms that there is not yet an appropriate evaluation methodology in the field and proceeded by defining this evaluation metrics:

Evaluation methodology
- *Accuracy*
- *Parsimony*
- *Recall*
- *Precision*
- *Production*
- *Learning Accuracy Results*

The research ended with these results: when using only tag collocation, the first step clustering algorithm can achieve a maximum F-measure of 59.7%4. The user collocation measure achieves a very similar result with a maximum F-measure of 59.1%5.

Factors influencing the effectiveness of performance measurement systems

A Knowledge Organization needs to measure its performance efficiently.

A successful Performance Measurement System (PMS) allows an organization to match its strategy with a measured and controlled performance. In this way, it allows us to determine if goals are being achieved but also identifies areas of improvement.

This paper explores the between multidimensional performance measures and the effectiveness of Performance Measurement Systems (PMS).

An effective PMS encourages the alignment of performance with desired goals, allows a more detail-oriented tracking of areas of improvement and it can help to build future strategies.

Furthermore, PMS has evolved from consisting only of financial measures, up to also including non-financial measures. This since outcome-centred measures have some disadvantages: first, it prevents managers from measuring employees' integral performance over strategic objectives; second, these measurements rarely indicate what needs to be fixed; third,

it can encourage short-term results at expense of long-term stability.

The **Balance Score Card** (BSC) allows to join financial with non-financial measures as it includes other areas as customer satisfaction, internal processes, and learning. At the same time, it enables the components of the organization to align to the firm's strategy and contribute to it.

Hypotheses Formulation
The authors formulate five hypotheses, which they evaluate then by means of a survey. The hypotheses are the following:

- *Hypothesis 1*: extensiveness of multidimensional performance measures is associated with effectiveness of PMS.
- *Hypothesis 2*: extensiveness of top management support is associated with effectiveness of PMS.
- *Hypothesis 3*: extensiveness of PMS-related training is associated with effectiveness of PMS.
- *Hypothesis 4*: extensiveness of employee participation in designing PMS is associated with effectiveness of PMS.

The evaluation of these hypotheses is conducted by means of a survey questionnaire of a random sample of 445 Australian Manufacturing business units, from which 118 questionnaires were used to carry out the data analysis.

The study finally finds that organizations have used PMS mainly to implement strategy and objectives but not so extensively to address concerns of the staff, for instance. This is an important area of improvement since results are achieved by focusing on both staff and performance related topics. As a matter of fact, the paper provides evidence on how staff-related outcomes serve as a boost for performance-related outcomes achievement.

Furthermore, the study finds evidence that financial measures are still being more extensively used on their own than coupled with non-financial measures. For this reason, the authors encourage firms to adopt more extensively non-financial oriented measures into their PMS.

Finally, authors provide evidence on how multidimensional performance measures along with top management support and training, displayed a significant association with PMS's effectiveness.

Conclusion

We understood that using modelling in a knowledge organization context improves not only the firm's working performance but also the human performance. The principal issues that a firm should address are the budget for the organizational capital and the knowledge application process. In addition, it is important to have an efficient information retrieval system to access the knowledge within the organization. We have seen that another modelling perspective is the need for a reliable performance measurement system to determine if goals are being achieved but also to identify areas of improvement.

Bibliography

[1] *Designing a maturity model for analyzing information and knowledge management in the public sector*; Jääskeläinen, A; Sillanpää, V; Helander, N; Leskelä, R; Haavisto, I; Laasonen, V; Torkki, P (2020)

[2] *Knowledge Management, Knowledge management system, and Organizational Performance: An Empirical Study*; Han, W; Y. Wang, Y (2012)

[3] *Using critical KM to address wicked problems, Knowledge Management Research & Practice*; John Dumay (2020); DOI: 10.1080/14778238.2020.1790310

[4] "*Knowledge management and hybridity of institutional logics in the public sector*."

Knowledge Management Research & Practice: 1-15; Laihonen, Harri and Petra Kokko (2020)

[5] *An Organizational Capital Decision Model for Knowledge-Intensive Organizations;* Januj A. Juneja, and A. D. Amar

[6] *Relating ontologies with a fuzzy information model*; Leite, MAA; Ricarte, ILM (2012)

[7] *Sense induction in folksonomies*; Andrews, P; Pane, J; Zaihrayeu, I

[8] *Effects of PMS process quality in construction firms*; Deng, F; Smyth, HJ; Anvuur, A (2013)

[9] *Factors influencing the effectiveness of performance measurement systems*; Tung, A; Baird, K; Schoch, H. (2011)

Ontology Model and Conceptual Mapping

Introduction

As we have seen in the Sankey Diagram of the bibliometric analysis of the Systematic Review, ontology became the more important and more used kind of knowledge based system during these years, so we decided to apply it.

To have a deeper analysis of our model and to better identify and understand underlying connections between entities, we extracted an ontological model based on our systematic review.

Model

This model consists of a representation of the 29 most important concepts cited in the systematic review as well as the properties found between them, by means of analysing data. To model our ontology, we made use of the software *Protégé* together with its plugins *OntoGraf, OWLViz* and *VOWL*.

According to our review, from the entities ontology [FIG. 6] it can be seen how Knowledge Management is divided in four main subclasses:

- Knowledge Organization
- Knowledge Creation
- Knowledge Transition
- Knowledge Application

It is evident how the last one particularly influences part of an enterprise performance, by using some "*Performance Measurement System*" to improve a firm "*Organizational Level Performance*".

As this could already be seen from the review, there are many other interesting underlying connections between classes that can be seen only by analysing the model:

Looking at the relationships ontology [FIG. 7), Information Retrieval System are a bridge between Data and a Knowledge Organization System; this systems directly influences performance and individual ability and since Information Retrieval Systems use in the specific public data, we can confidently say that implementing Data Lakes could improve both Knowledge Management and Individual Performance in any enterprise.

Another important concept that can be extracted is how Human Capital generates Profit in an enterprise. As we can see, enterprises are strongly connected to data, and since public data improves knowledge management, we can claim that applying knowledge information to human capital would directly increase both profit and performance.

Referring to "*Using critical KM to address wicked problems*" by John Dumay we mark the importance of how KM should be used by enterprises to address wicked situations that taint our modern jobs world. Human Capital is exactly this: it is something that goes beyond economics, that aims to improve the working conditions in the whole system.
Knowledge Transfer between enterprises could generate public data, and if on one hand sharing productivity and performance measures could help contenders on the market, sharing working conditions and human resources practices would for sure be the first step towards endorsing much bigger problems.

Our Systematic Review strongly supports this point of view: Information Retrieval is core in a Knowledge Organisation ecosystem and the amount of links it has in the ontology model are a visual proof of this concept; investing collective strength and resources in this field could surely improve

the lack of documentation on how Knowledge Management should also be a tool to improve the modelling of human actions.

Concluding, Knowledge Management is not only useful towards the improvement of capital, profits and performance, but is key to the connections between every single step of the chain that allows success in an enterprise.

Ontology Graphs

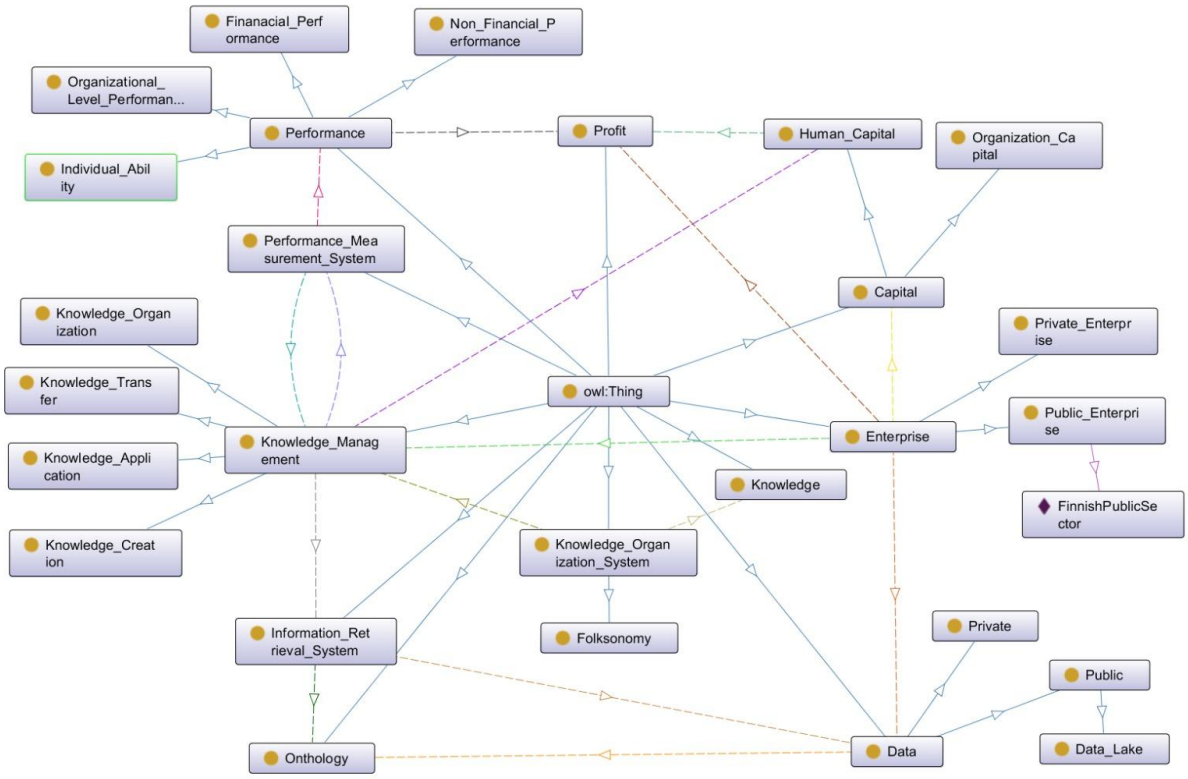

FIG. 6: Entities Ontology. We can visualize the links between different concepts without caring about what kind of links they are.

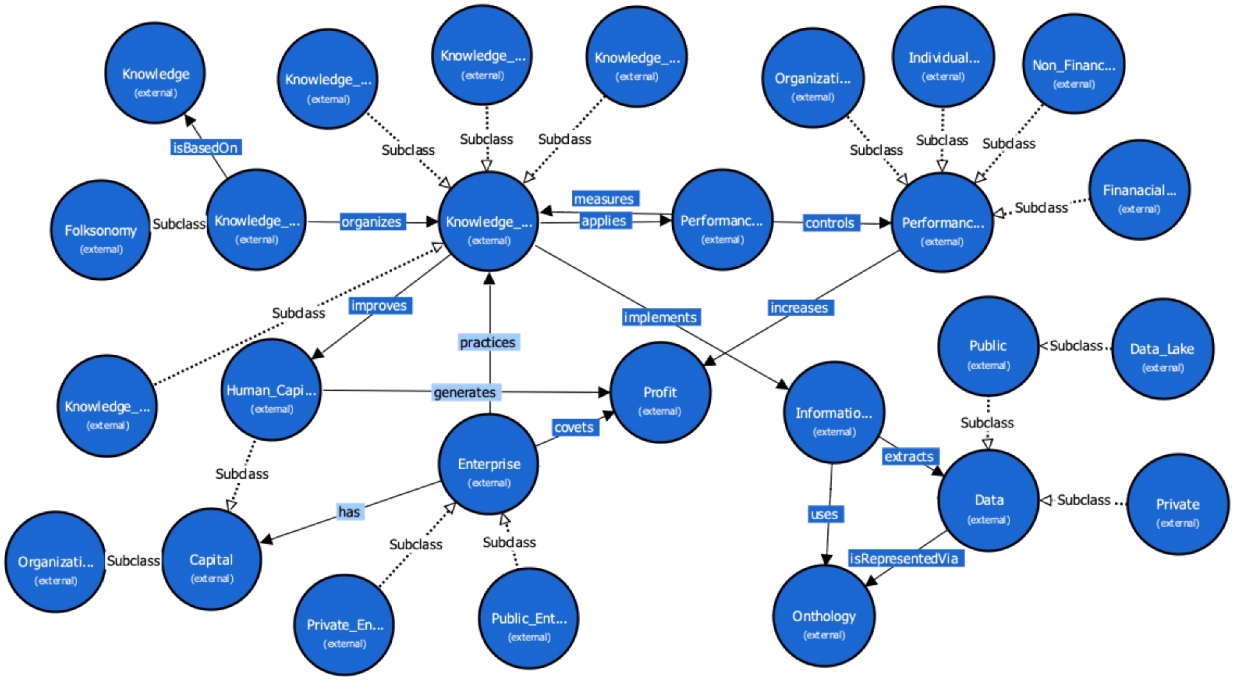

FIG. 7: Relationships Ontology. We can visualise the different types of existing relationships, in order to learn new things about our model.

Empirical Analysis

Introduction

After the Systematic Review and the definition of an Ontology model, we decided to perform an empirical analysis on a real dataset. We hope to gather some relevant information about the knowledge organizations and how they can affect, or they are affected by, human behaviour.

Description of Chosen Method

To carry out our Empirical Analysis, we chose to use the *Visual Analytics Environment (VAE)*. This decision was taken mainly for two reasons:

1. The powerful analyses that can be done with the tool, despite its ease of use.
2. The data we chose to examine is best suitable to be analysed by a Visual Analytics tool rather than a Cost Benefit Analysis or a Probabilistic Graphical Model.

Description of Chosen Data

We know that the data quality is defined as “fitness for use” with respect to users' needs. It has seven dimensions:

1. **Relevance**. Degree to which data serves to address their purposes.
2. **Accuracy**. How the data correctly describes the features they are designed to measure.
3. **Credibility**. Confidence of the users in the data products and trust in the objectivity of the data.
4. **Timeliness**. Length of time between their availability and the phenomenon they describe.
5. **Accessibility**. How readily the data can be located and accessed.
6. **Interpretability**. The ease with which the user may understand and properly use and analyze the data.
7. **Coherence**. The degree to which they are logically connected and mutually consistent.

The dataset we decided to analyze is the *Digital Economy and Society Index* (DESI), which perfectly suits our purposes in terms of timeliness and accessibility, in fact since 2014 not only evaluates Europe’s digital performance but monitors the digital progress of EU countries by using a vast number of indicators and everyone is allowed to access the data being even helped in analyzing it by multiple tools that the very European Commission propose in its dataset.

The indicators taken into account in our analysis are indeed an optimal option since they are very precisely described by our data, making the accuracy of it particularly high. These indexes seek to help EU members to identify areas in which priority action must be taken, by directing public policy. Its credibility is given by multiple factors as well as the importance of the institution of the European Commission and its history.

Relevance of the DESI dataset becomes crystal clear as it analyzes simultaneously the two major components of this project: knowledge organization and human action in different - but similar enough - societies. Knowledge organization as it measures internet usage, digital technologies for businesses, digital public services for citizens and businesses, among others. Whereas with respect to Human Action, it tracks digital skills, eCommerce, eGovernment, eHealth, among others.

Even more, the vast number of indicators provides us flexibility to create more heterogenous and robust analysis by using the VAE highlighting the great coherence that runs between the indicators and how

they are logically connected (see Fig. 6 the ontology) and mutually consistent.

DESI also provides information based on 5 key areas: human capital, connectivity, integration of digital technology, digital public services and Research and Development in Information and Communication Technology (ICT). This extensive information enables us to create a more integral analysis.

Finally, the DESI addresses several components of our previously exposed ontology, such as: capital, knowledge, information retrieval, human capital, private and public enterprises, among others and the great literature that can be found on this topics allows a great interpretability to the user to properly analyse all other features.

Analysis

Among the vast number of indicators of the DESI dataset, we decided to focus our attention to the following variables:

- **At least basic digital skills**. Percentage of all individuals (aged 16-74) with 'basic' or 'above basic' digital skills in each of the following four dimensions: information, communication, problem solving and software for content creation (as measured by the number of activities carried out during the previous 3 months).
- **ICT Specialists**. Percentage of total employed ICT specialists. It includes jobs like ICT service managers, ICT professionals, ICT technicians, ICT installers and servicers.
- **Enterprises providing ICT training.** Percentage of enterprises who provided training in ICT to their personnel.
- **SMEs with at least a basic level of digital intensity.** The digital intensity score is based on counting how many out of 12 selected technologies are used by enterprises. A basic level requires usage of at least 4 technologies. The indicator measures the percentage of SMEs (10-249 persons employed) which satisfies this requirement.
- **Electronic Information Sharing.** Percentage of enterprises who have in use an ERP (enterprise resource planning) software package to share information between different functional areas (e.g. accounting, planning, production, marketing).
- **Big data**. Percentage of enterprises which analyse big data from any data source.
- **e-Government users**. Percentage of all individuals (aged 16-74) who used the internet, in the last 12 months, for interaction with public authorities.
- **Digital public services for citizens**. The score of the country, from 0 to 100, which indicates the share of administrative steps that can be done online for major life events (birth of a child, new residence, etc.) for citizens.
- **Open data**. Percentage of the maximum open data score which measures to what extent countries have an open data policy in place, the estimated political, social and economic impact of open data and the characteristics (functionalities, data availability and usage) of the national data portal.

We need to give the definition of what we consider under the name 'enterprise' according to the DESI. We consider the enterprises with 10 or more persons employed, from all manufacturing and service sectors, excluding the financial sector.

We collected data from 2019 to 2021 for all the variables, except for digital public service for citizens, open data and SMEs with at least a basic level of digital intensity for which we have only data referring to 2021.

We decided to carry out the following analyses through VAE: correlation analysis, monodimensional analysis, geographical analysis, ratios analysis, multidimensional analysis and anomalies check (Big Data, Basic Digital Skills and eGovernment users).

Correlation Analysis

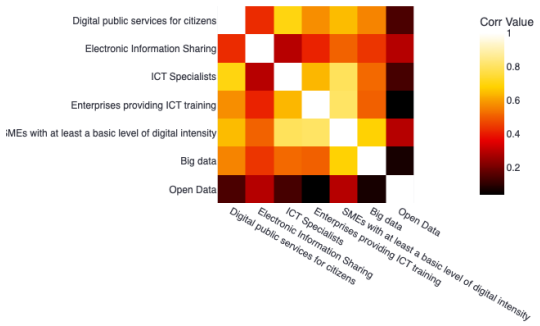

FIG.8: Correlation Analysis made from DESI Indexes by using Visual Analytics Environment (VAE)

Laihonen, H & Kokko, P (2020) manifest how a hybrid approach of public data coupled with more involved organisations can positively impact performance.

A correlation analysis over some of the DESI index data allows us to see how the digital intensity of Small and Medium Enterprises is correlated with the use of Big Data, the proportion of enterprises that provide ICT training and the digital public services for citizens.

Therefore, despite not having indicators of performance, we can appreciate the positive relationship between the digital intensity of organisations and the digitalization of public services offered to citizens. In this way, we would expect that a greater intensity in digitalization of firms will improve the public services given to citizens, which finally will have a positive impact on performance of economic actors, as better services are associated with greater achievements.

Mono-dimensional Analysis

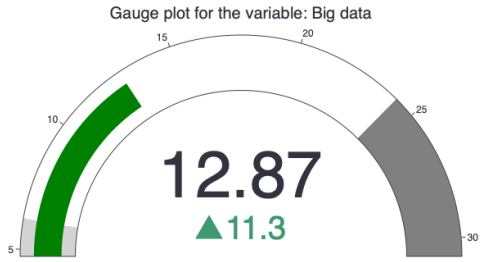

FIG.9: Monodimensional Analysis of Big Data variable from DESI Indexes made by using Visual Analytics Environment (VAE)

The Big Data variable from the DESI Indexes makes reference to the relative number of enterprises that analyse big data from any data source. The mono-dimensional analysis allows us to see that on average 12.87% of the enterprises, from all the European countries studied in the DESI Indexes, are analysing big data.

Then, the 95th percentile of firms analysing big data is slightly below 25%. This certainly means there is still a big room for improvement since relatively just a few enterprises are analysing big data sources. The latter makes sense with what Drucker, P & Maciariello, A (2008) expose about how data analysis has a key role in the decision making of a good manager.

Hence, it is possible to conclude that if a bigger number of firms shift to a more intense big data analysis, this could translate into better decisions taken by managers and therefore a better performance.

In this way, it is possible to visualise the deep relationship between management in a knowledge organisation, human modelling and performance.

Geographical Analysis

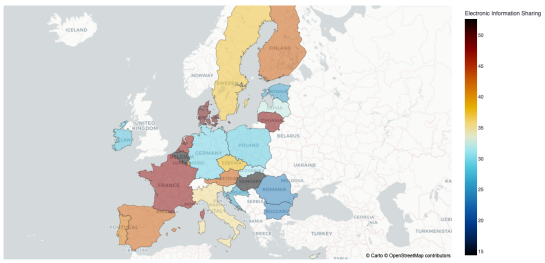

FIG.10: Geographical Analysis of Electronic Information Sharing variable from DESI Indexes made by using Visual Analytics Environment (VAE)

The Electronic Information Sharing (EIS) variable makes reference to the percentage of enterprises which have a software for sharing information between different functional areas.

The Geographical Analysis does not necessarily allow us to spot patterns by geographical localization regarding EIS. Countries like Belgium, Denmark, France, Netherlands and Lithuania appear to be at the top of this indicator with relative numbers oscillating between 40% and slightly above 50%. The second group is made up of Finland, Sweden, Portugal, Spain, Austria and Italy, whose EIS oscillates between 35% and 45%. Lastly the rest of the analysed countries have an EIS approximately below 30%.

Ratio analysis

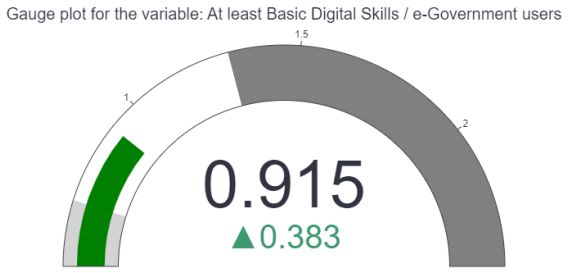

FIG.11: Gauge plot of the Ratio Analysis made from DESI Indexes by using Visual Analytics Environment (VAE)

We decided to look at the ratio between the percentage of all individuals with at least basic digital skills and the percentage of individuals that, in the last 12 months, used the internet to interact with the public authorities.

When this ratio is greater than 1 it means that the individuals know how to use technology but they do not use it to interact with the public authorities. While if the ratio is smaller than 1, we can say that a percentage of individuals uses the internet to interact with the government but it does not have basic digital skills.

We have a mean value of 0.915, which means that in many countries there are people with no basic digital skills that use the internet to interact with the public environment. We can say that in these countries the state encourages the use of the technology more than the others.

Below, we can see the violin plot which shows us that only a few countries have a ratio much greater than one and the distribution is concentrated under 1. As we could see from the dark grey bar in the gauge plot the values greater than 1 are located over the 95th percentile, so we can consider them as outliers.

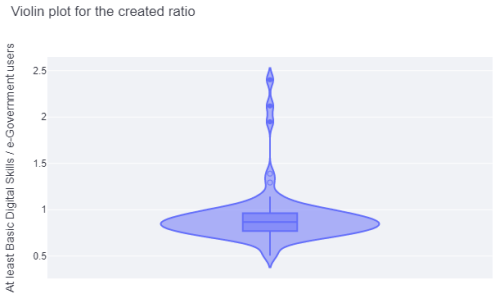

FIG.12: Violin plot of the Ratio Analysis made from DESI Indexes by using Visual Analytics Environment (VAE)

Multidimensional analysis

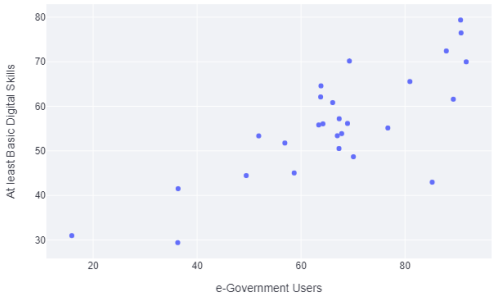

FIG.13: Multidimensional Analysis made from DESI Indexes by using Visual Analytics Environment (VAE)

In this multidimensional analysis we have a scatterplot of the e-Government users and the individuals with at least Basic Digital Skills. As we can see from the plot, there is a positive correlation between the two variables. So, a high number of people with at least basic digital skills corresponds to a high number of e-Government users.

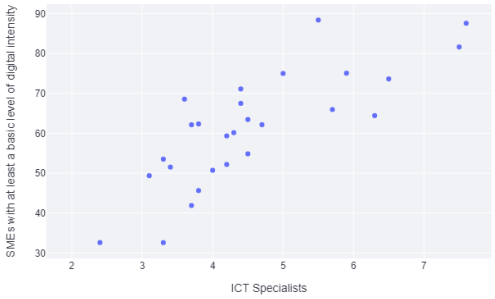

FIG.14: Multidimensional Analysis made from DESI Indexes by using Visual Analytics Environment (VAE)

Here, we have a scatterplot between the number of ICT specialists employed in the firms and the SMEs with at least a basic level of digital intensity. We can see that, obviously, the large number of ICT specialists employed in a country corresponds to higher values of SMEs that apply technology information inside their decisional processes.

Anomalies check

The anomalies check from the VAE tool allows us to see if we have some outliers inside the distributions of our variables.

We focused our attention on the e-Government users. The result obtained tells us we have only left outliers. In particular, we have 6 strong left outliers (Romania, Italy in 2019 and 2020 and Bulgaria in 2019) and 5 weak left outliers (Bulgaria in 2020 and 2021, Croatia in 2020, Italy in 2021 and Poland in 2019). The lowest result has been registered in Romania where in 2021 only less than 16% of the population used the internet to access public infrastructure. Anyway, we can highlight that also in these countries the percentage is increasing during the years.

Country	*Year*	*e-Government Users*
Bulgaria	2019	33.0
Bulgaria	2020	36.03
Bulgaria	2021	36.28
Croatia	2020	41.25
Italy	2019	31.73
Italy	2020	29.88
Italy	2021	36.35
Poland	2019	44.78
Romania	2019	12.09
Romania	2020	14.61
Romania	2021	15.89

Then, we considered the variable about the percentage of enterprises which analyse big data from any data source. In this case, we have only weak left outliers, so countries with a very low level with respect to the others.

These countries are Cyprus, Romania and Slovakia.

Country	*Year*	*Big data*
Cyprus	2019	4.73
Cyprus	2020	4.73
Romania	2021	5.1
Slovakia	2021	5.6

In our opinion, in recent years the usage of big data and the digitalization of the public environment are crucial for the development of a country and of a firm. Under this perspective, these left outlier values from the previously cited countries absolutely need to be improved.

Conclusions

After the creation of the ontological model and the analysis of a real dataset about the digital performance of public and private institutions, we can delineate some conclusions about the topic of our project.

We understood that using modelling in a knowledge organization context improves not only a firm's working performance but also the human performance.

From the ontology model, we learn the importance of good knowledge management and knowledge transfer.

Knowledge Management is not only useful towards the improvement of capital, profits and performance, but is key to the connections between every single step of the chain that allows success in an enterprise.

Moreover, Knowledge Transfer is fundamental to generate public data and so to solve more complex problems that can’t be solved from an enterprise on its own.

Analyzing the last year's trends, we noticed that the implementation of artificial intelligence mechanisms to knowledge management led to the development of efficient knowledge organization.

Technology has a central role in this evolution. Looking at the real data from the DESI dataset we could have seen the state of the art in the digitalization of both firms and public organizations.

We have noticed that the percentage of enterprises that implemented big data into their processes is still too low. We know from the book Management, written by Drucker & Maciariello, that the data are really important into the decision making process of a firm. They can be really helpful to the manager to take the right directions for its firm.

Also the public governments need to improve their digitalization and their policies on data. We have seen as a good example the Finnish model where data lakes were used to improve performances in many contexts of society.

Unfortunately, Italy is one of the countries that needs more improvement in this field. We have seen that its level of users that interact with the public authorities through the internet is one of the lowest in Europe. But, from 2020 to 2021 we have observed a considerable increase in these values.

In our opinion, the Covid-19 pandemic accelerated the process of digitalization in our country both for the public sector and the private enterprises, this means that maybe we are going in the right way.

Finally, our model confirms that they should use big data and knowledge management to model human behaviour in order to improve their performances.

OTHER REFERENCES

- *The Organization of Information, Daniel N. Joudrey, Arlene G. Taylor, 2017.*
- *Introduction to the course and building blocks, Daraio C., Slides of the course Quantitative Models for Economic Analysis and Management, 2022.*
- *An introduction to Modelling and Implementation, Daraio C., Slides of the course Quantitative Models for Economic Analysis and Management, 2022.*
- *Management - Revised Edition, Peter F. Drucker, Joseph A. Maciariello, 2008.*
- *Data I: Nature, collection, semantic modelling, and analysis of big data and little data within organizations and socio-economic systems, Daraio C., Slides of the course Quantitative Models for Economic Analysis and Management, 2022.*
- *Seminar: Ontology tools, Di Leo S., Slides of the course Quantitative Models for Economic Analysis and Management, 2022.*
- *Digital Economy and Society Index (DESI), Digital Agenda Data EU, Provided by the European Commision, 2021.*

APPENDIX

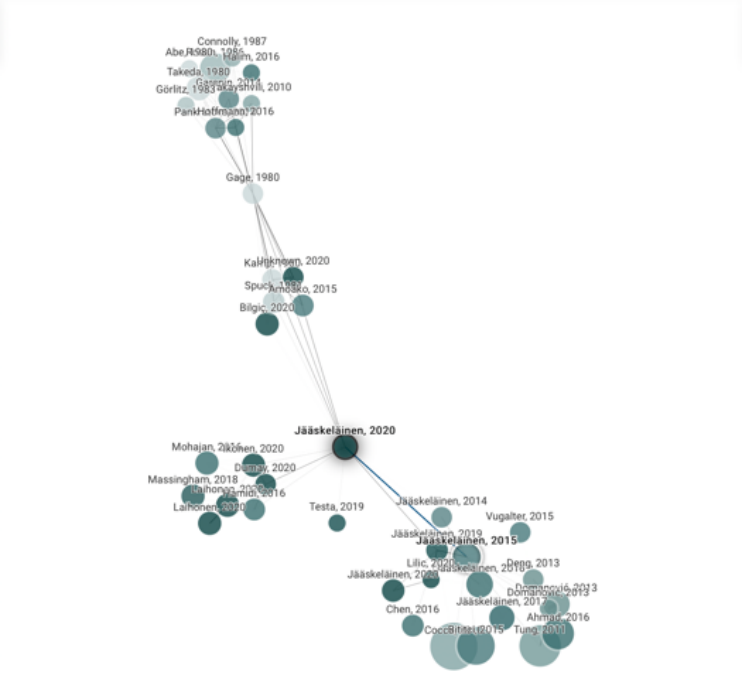

FIG. 1: Connected papers of "*Designing a maturity model for analyzing information and knowledge management in the public sector*" by Aki Jääskeläinen.

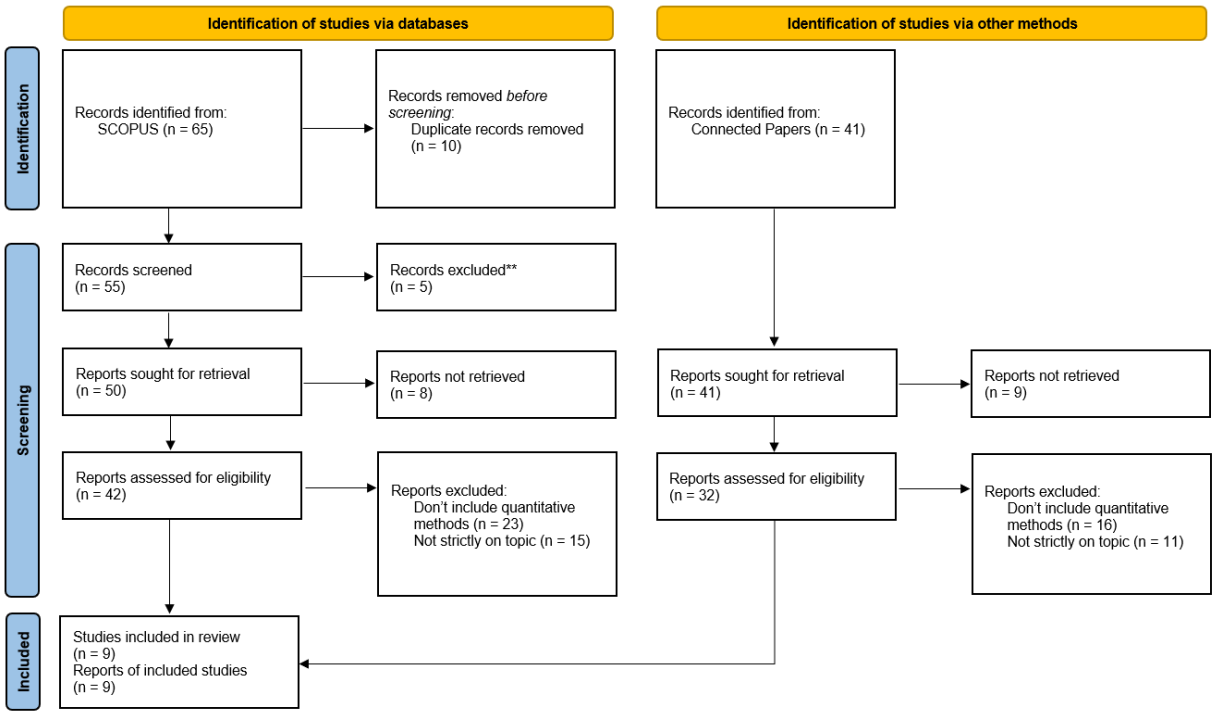

FIG. 2: PRISMA 2020 flow diagram for the Systematic Review "*The Performance of Knowledge Organizations and Modelling Human Action*".

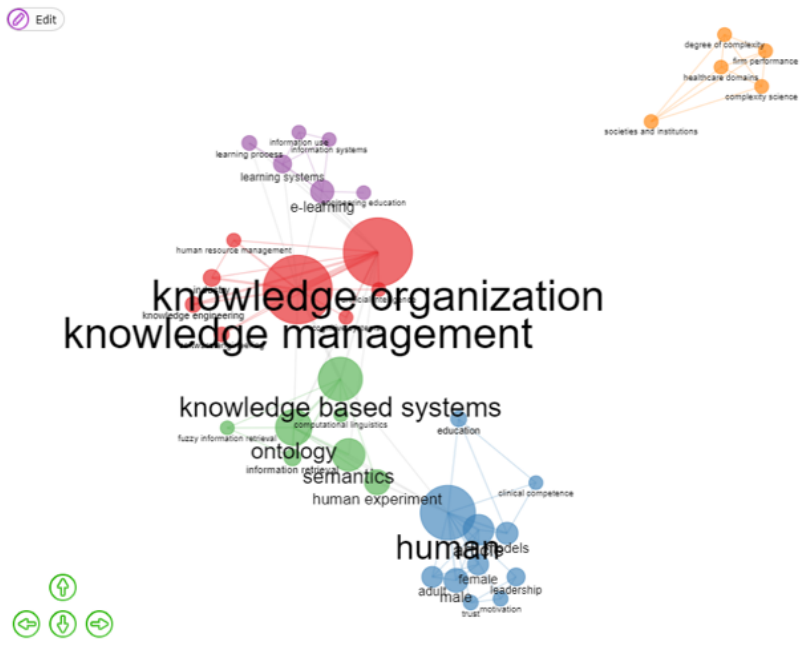

FIG. 3: Through Bibliometrix: Co-occurrence network of the 65 articles found in SCOPUS. The fields are the keywords plus. The most used keywords are knowledge management, knowledge organization and human.

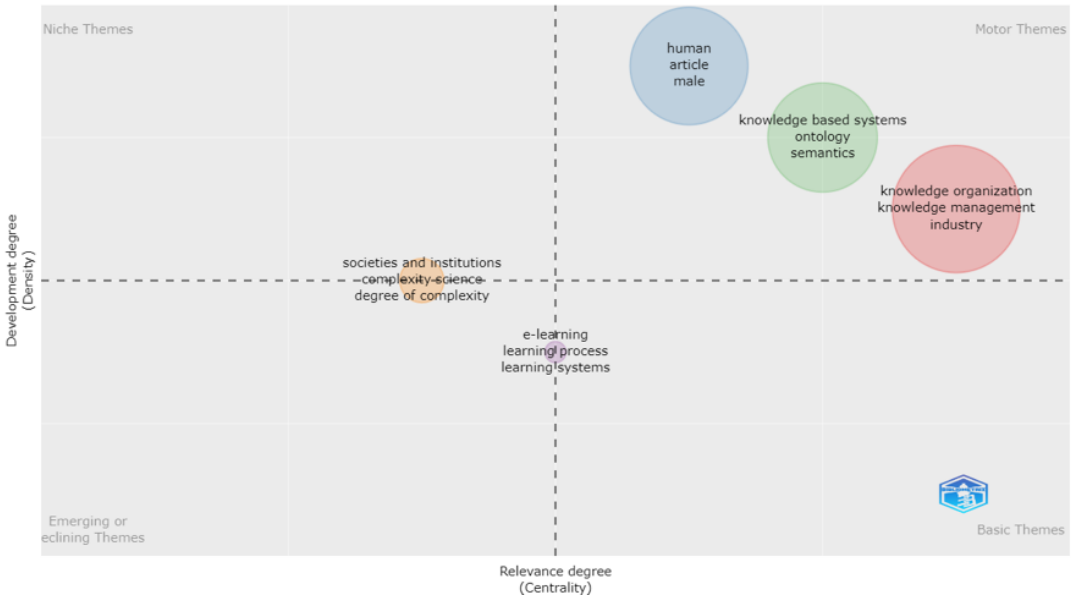

FIG. 4: Through Bibliometrix: Thematic Map of the 65 articles found on SCOPUS. We can see that the motor themes of our research are human actions, knowledge-based systems (ontology, semantics) and knowledge organization and management.

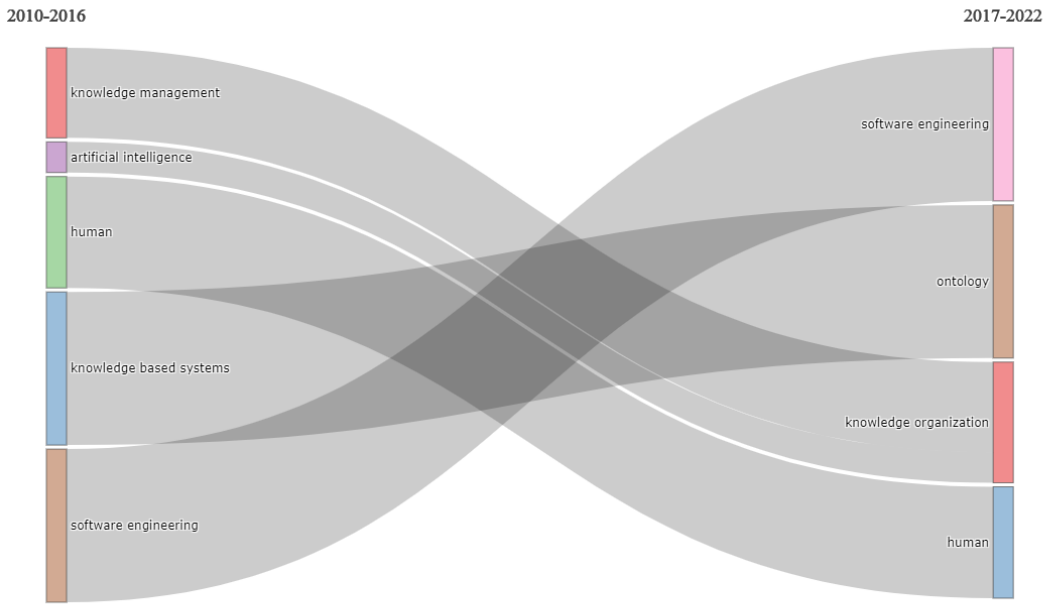

FIG. 5: Through Bibliometrix: Thematic Evolution Map of the 65 articles found on SCOPUS. We can see some interesting developments in the papers. Knowledge management and artificial intelligence topics moved to knowledge organization and the knowledge-based systems focused on ontology.