The University of Texas at Dallas

School of Economic, Political and Policy Sciences

EPPS 6302: Methods of Data Collection and Production

Prof. Dr. Karl Ho

Analyzing the Scientific Evolution of Artificial Intelligence in Educational Field Using Text Mining

Student name: Federico Ferrero

Fall, 2019

EPPS 6302 Methods of Data Collection and Production

Final Report

Analyzing the Scientific Evolution of Artificial Intelligence in Educational Field Using Text Mining

Federico Ferrero

Introduction

Nowadays, the already known Artificial Intelligence techniques take a new boost based on the computation capacity increment and the accumulation and constant updating of large quantities of data (Danaher et al., 2017; Holmes, Bialik, and Fadel, 2019).

Within this "datafication" context (Breiter, 2016; Selwyn, 2015; Van Dijck, 2014) these systems use algorithms to assess situations and make decisions covering a large variety of areas that have an impact at private individual level. In particular, these automated systems are strongly present in business, since it was there where the so-called "predictive risk analytics" (Siegel, 2016) first found applicability. However, algorithms are used more and more to make decisions in the area of health, in justice and in the organisation of prisons, in urban design and its mapping, in government and bureaucratic systems (Batty, 2013; Jee and Kim, 2013; Kim, Trimi, Chung, 2014; Khel, Guo and Kessler, 2017) and, recently, in the educational/pedagogical contexts. In this field, algorithms are usually used to predict performances, choose students and assess teachers, or to develop "Intelligent Mentoring Systems" or "Adaptive Learning Systems" to recommend lessons and contents to pupils, amongst others (Aleven, et al. 2015; Baker, 2016; Daniel, 2017; Sclater, Peasgood and Mullan, 2016; Simpson, 2006; Williamson, 2017).

Having said that, the majority of Artificial Intelligence experiences seem to be isolated and little research has been carried out with meta-analytical intentions over the specific educational research field. Therefore, this project proposes the analysis of Artificial Intelligence scientific production in the educational field during the last 30 years considering not just its conceptual structure but also its scientific evolution.

In this context, this research uses text mining technics applied over bibliographic material and proposes to answer two specific questions: a) what are the main topics of Artificial Intelligence investigated in the field of educational research during the last three decades?; and b) which is its thematic evolution during this period?

Methodology

This exploratory study performs a Systematic Literature Review (SLR) whereby are used systematic and explicit methods to find, select, and critically assess the relevant research starting from a question/problem, with the aim of mapping the field of knowledge in the current moment (Meca, 2010; Okoli and Schabram, 2010).

Doing so, it is proposed the work with SciMAT, a science mapping analysis software tool developed by Cobo, López-Herrera, Herrera-Viedma, and Herrera (Cobo et al. 2011; 2012; Martinez et al. 2015) from the University of Granada, Spain. This software application is a bibliometric science mapping tool based on co-word analysis and h-index. Besides, it works in a longitudinal framework in order to detect the different themes treated by the research field across the given time periods.

The following steps were considered to carry out the analysis:

1) **Collection of the raw data**

After different alternatives of searches, it was decided to examine the Web of Science Core Collection (ISI WoS) according to the following criteria:

> *(TS=((artificial intelligence OR success student algorithm OR intelligent tutoring systems OR big data) AND education)) AND DOCUMENT TYPES: (Article)*
> *Refined by: WEB OF SCIENCE CATEGORIES: (EDUCATION EDUCATIONAL RESEARCH OR PSYCHOLOGY DEVELOPMENTAL OR COMPUTER SCIENCE INTERDISCIPLINARY APPLICATIONS OR EDUCATION SCIENTIFIC DISCIPLINES OR COMPUTER SCIENCE INFORMATION SYSTEMS OR INFORMATION SCIENCE LIBRARY SCIENCE OR PSYCHOLOGY MULTIDISCIPLINARY OR SOCIOLOGY OR COMPUTER SCIENCE ARTIFICIAL INTELLIGENCE OR PSYCHOLOGY EDUCATIONAL OR PSYCHOLOGY EXPERIMENTAL OR SOCIAL SCIENCES INTERDISCIPLINARY OR PSYCHOLOGY SOCIAL OR COMPUTER SCIENCE THEORY METHODS OR COMMUNICATION OR EDUCATION SPECIAL OR PSYCHOLOGY APPLIED)*
> *Timespan: All years. Indexes: SSCI, A&HCI, CPCI-SSH, BKCI-SSH.*

Also, it was established a filter by categories linked to education because many of the papers corresponded to fields other than educational/pedagogical. This search yielded a final result of N=770 documents downloaded by September 2019.

At the same time, as Artificial Intelligence is a topic of incipient development in the educational field (the first record that satisfies the search carried out is from 1984 and the total number of papers amounts to only 770); it was determined to partition periods every 5 years to generate a more interesting evolution diagram.

2) **Normalization of keywords**

The keyword was defined as the item to analyze. This includes authors keywords, journals keywords, and indexing keywords presented in the selected documents.

For its normalization, keywords were chosen in the plural rather than in the singular form and with a hyphen rather than without a hyphen. In this way, the keywords were joined in groups automatically. After this, a manual process of location was conducted with the keywords that did not were classified by the software.

3) **Co-occurrence frequencies of keywords and similarities between items**

First, a calculation was conducted by counting the number of documents in which the two keywords appear together. Then, the Equivalence Index was determined as SciMAT considers it as the most appropriate measure for normalizing co-occurrence frequencies. The interpretation of this value is between 0 and 1: when the keywords always appear together, the Equivalence Index equals 1 and when they are never associated, it equals 0.

4) Clustering

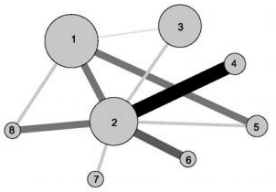
The next step in the process was the clustering. Through this technique, it was possible to locate subgroups of keywords that are solidly linked and that correspond to centers of interest in research. In this case, SciMAT works with Simple Centers Algorithm because it returns automatically labeled clusters.

5) Analysis and interpretation

In addition to Performance Analysis measures (such as the number of documents, authors, journals, received citations, h-index or measure of scientific research impact, etc.), four types of diagrams were considered:

1) Thematic diagram

This graphic consists in a group of nodes that represent different keywords connected. This group together is called a “theme”. The node size is proportional to the number of documents corresponding to each keyword and the thickness of lines between nodes are proportional to the Equivalence Index.

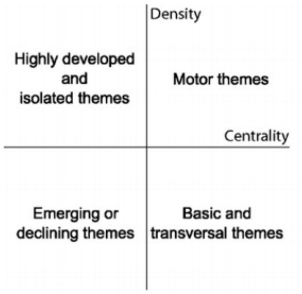


Example of Thematic diagram extracted from Cobo, López-Herrera, Herrera-Viedma and Herrera (2011)

2) Strategic diagram

This graphic locates “themes” according to two parameters: their centrality and density measures. The centrality measures the degree of interaction of a network with other networks, which means a measure of the importance of a theme in the development of the entire research field analyzed. The density measures the internal strength of the network, that is the strength of internal ties among all keywords describing the research theme. For this reason, this parameter can be understood as a measure of the theme’s internal development.

It is organized in four quadrants: in the lower-left sector are located the emerging or disappearing themes (with low centrality and density); in the upper-left quadrant are positioned the highly developed and isolated themes (with high density and low centrality); in the lower-right quadrant are the basic and transversal themes (high centrality and low density); and in the upper-right sector the motor themes are placed (with high density and centrality which means that they are important themes, connected with other themes, and developed internally as well).


Density
Highly developed and isolated themes
Motor themes
Centrality
Emerging or declining themes
Basic and transversal themes

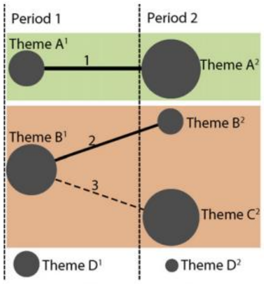
Structure of Strategic diagram extracted from Cobo, López-Herrera, Herrera-Viedma and Herrera (2011)

3) Thematic evolution diagram

This diagram shows the thematic evolution of the research field under study. The periods are organized vertically and in each one of them, different themes (spheres/nodes) are located.

Some themes are linked with solid lines showing thematic continuity. This solid line can include two different situations: both themes have the same name or the name of one of them is part of the other one. Also, two themes can be linked with a dotted line which means that they share elements that are not the name of the themes. This continuity is less strong than the solid-line one.

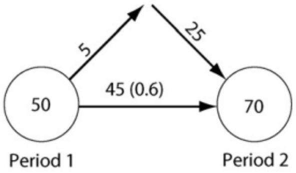
The thickness of the lines/edges is proportional to the Inclusion Index, and the volume of the nodes is proportional to the number of published documents associated with each theme.



Example of Thematic Evolution diagram extracted from Cobo, López-Herrera, Herrera-Viedma and Herrera (2011)

4) Stability between periods diagram

This diagram shows stabilities measures across the consecutive periods: circles represent the periods and numbers are the quantities of keywords in each one. Horizontal arrows are the keywords shared by two consecutive periods and the number in parenthesis indicates the Similarity Index. Also, the arrows ponting up indicate the number of outgoing keywords and the arrows pointing down the number of new keywords entering each period.



Example of Stability diagram extracted from Cobo, López-Herrera, Herrera-Viedma and Herrera (2011)

Results and analysis

Preliminary analysis

According to Figure 1, the explosion of documents on Artificial Intelligence published in the education field has been recorded since 2008 with sustained growth until the date of downloading the database (September 2019). From the geographical point of view (Figure 2) the main publisher countries were United States (n=205), United Kingdom (n=72), China (n=40), and Spain (n=40).

Figure 1

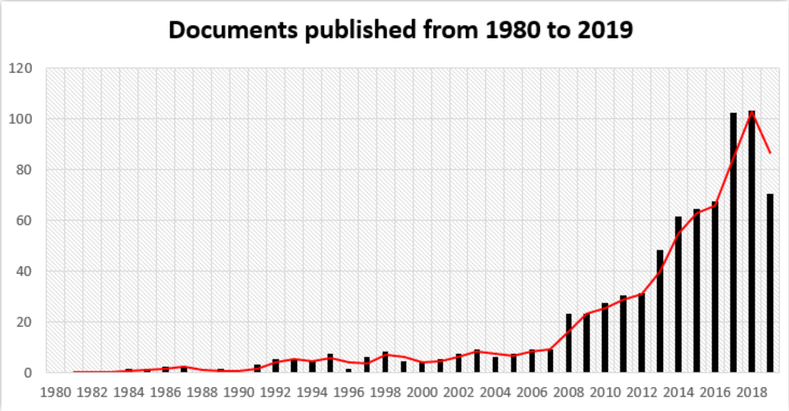
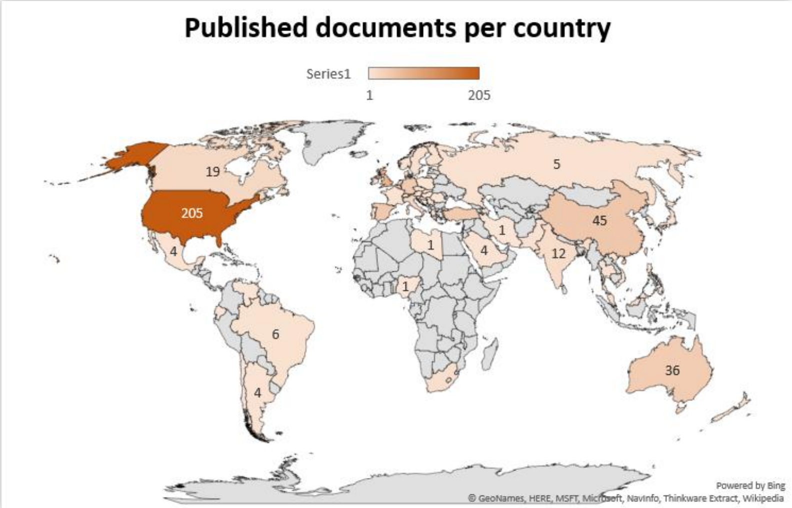


Figure 2


Published documents per country
Series1
1
205
19
205
4
6
4
5
45
1
4
1
12
1
36
Powered by Bing
© GeoNames, HERE, MSFT, Microsoft, NavInfo, Thinkware Extract, Wikipedia

Regarding the predominant publications (Figure 3), despite having a great dispersion, the main journals and magazines are Computers & Education and Computers in Human Behavior, two recognized means of dissemination of scientific works in the field. At the same time, a preliminary analysis with words recurrences carried out with Wordle on the Abstracts of the publications considered (Figure 4), allows us to observe that the concern moves from the Tutoring Systems (decade of the 90s) to typically pedagogical topics such as learning and students (decade 2000) towards discussions about learning and use of data (during the last decade).

Figure 3

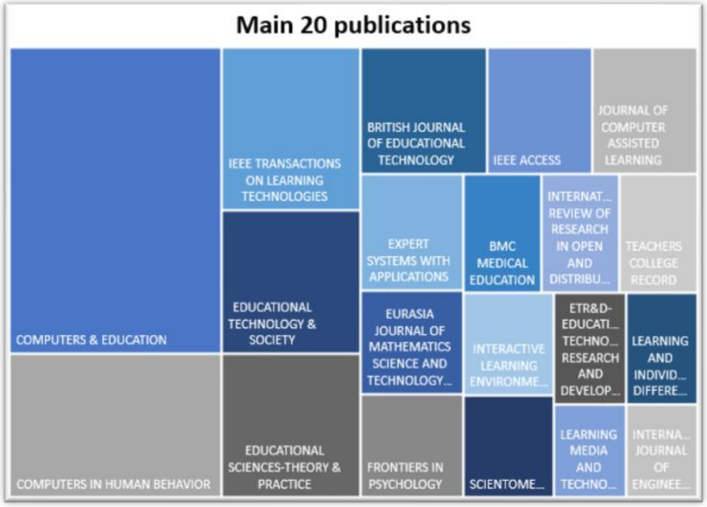
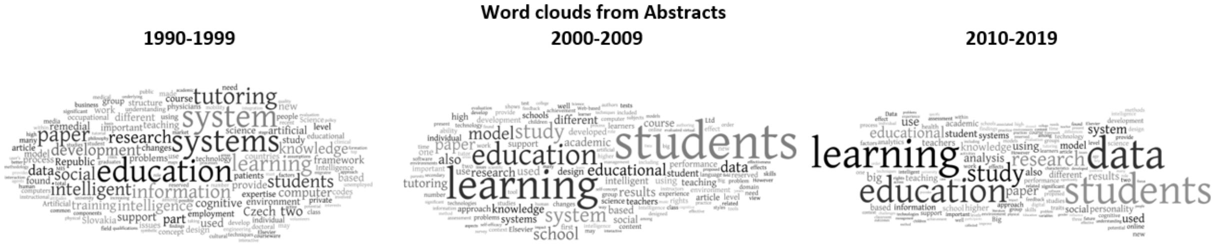


Figure 4



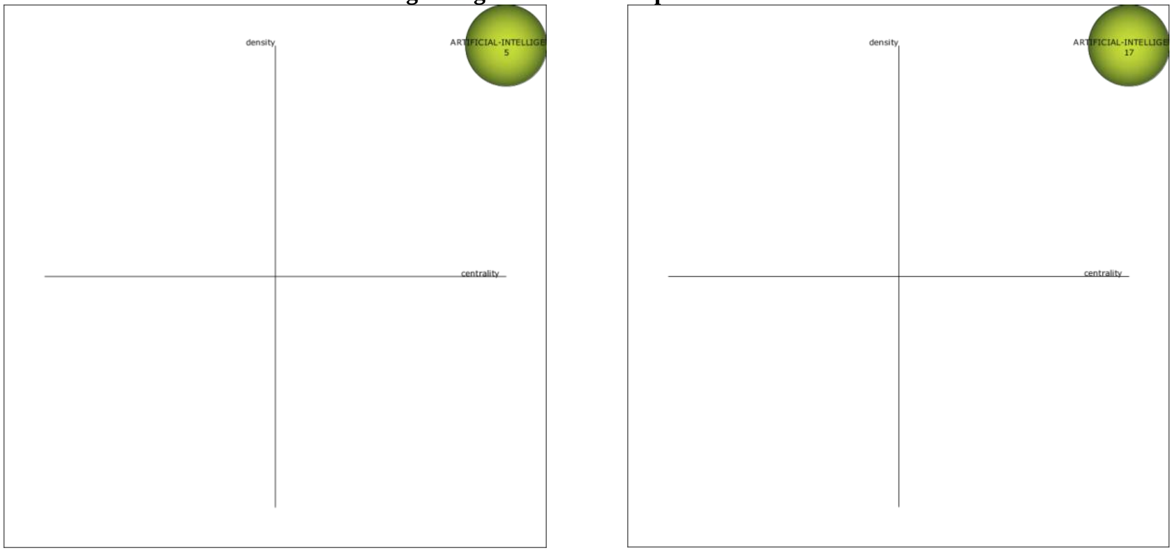
Strategic Diagrams

In order to analyze the most prominent themes in each predefined period, two strategic diagrams are presented for each one. On the one hand, a strategic diagram in which the volume of the sphere is proportional to the number of published core documents. On the other hand, a strategic diagram in which the volume of the node is proportional to the number of citations received for each theme.

The first analyzed period that presents a complete strategic diagram is 1991-1995 (Figure 5 and Table 1) where the theme “ARTIFICIAL-INTELLIGENCE” appears as a motor theme with just 5 documents published, 17 citations, and with little impact (h-index score of 2).

Figure 5

Strategic diagrams for the subperiod 1991-1995



Strategic diagram based on the number of published documents

Strategic diagram based on the number of citations received

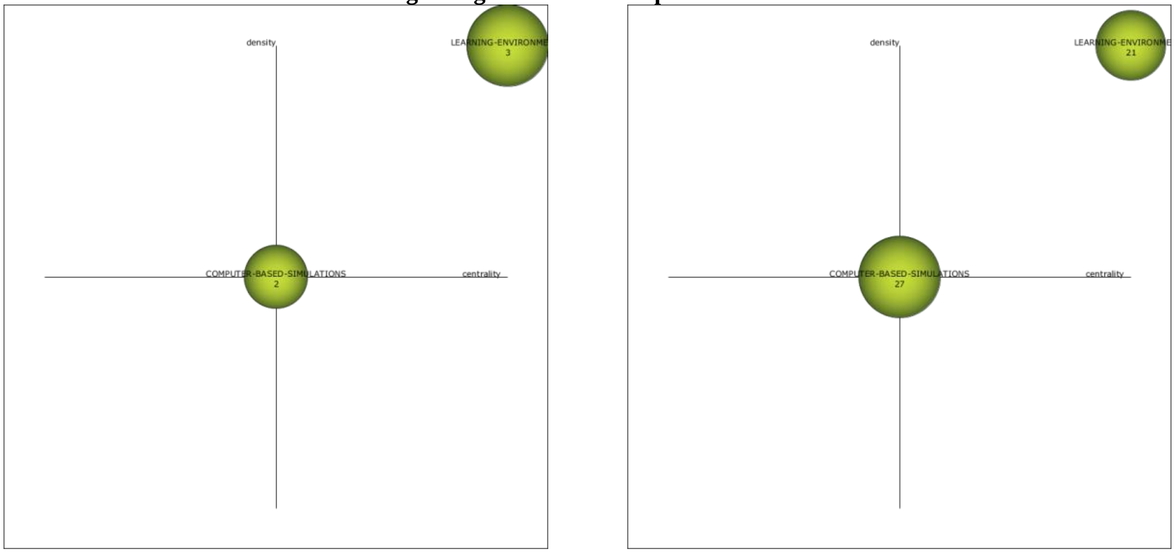
Table 1

PERIOD	Theme Name	Number of Documents	h-Index	Number of citations
1991-1995	ARTIFICIAL-INTELLIGENCE	5	2	17
1996-2000	LEARNING-ENVIRONMENTS	3	3	21
	COMPUTER-BASED-SIMULATIONS	2	2	27
2001-2005	DISTANCE-EDUCATION	9	8	528
	MODELING	2	2	122

The next period (1996-2000) (Figure 6 and Table 1) is slightly diversified with two new themes: “LEARNING-ENVIRONMENTS”, clearly a motor theme, and “COMPUTER-BASED-SIMULATIONS” with a considerable number of citations. Something interesting to note is that there is a subdivision of the previous theme into two topics that correspond to the two main disciplines involved in the field: Education Sciences and Computer Sciences, respectively.

Figure 6

Strategic diagrams for the subperiod 1996-2000



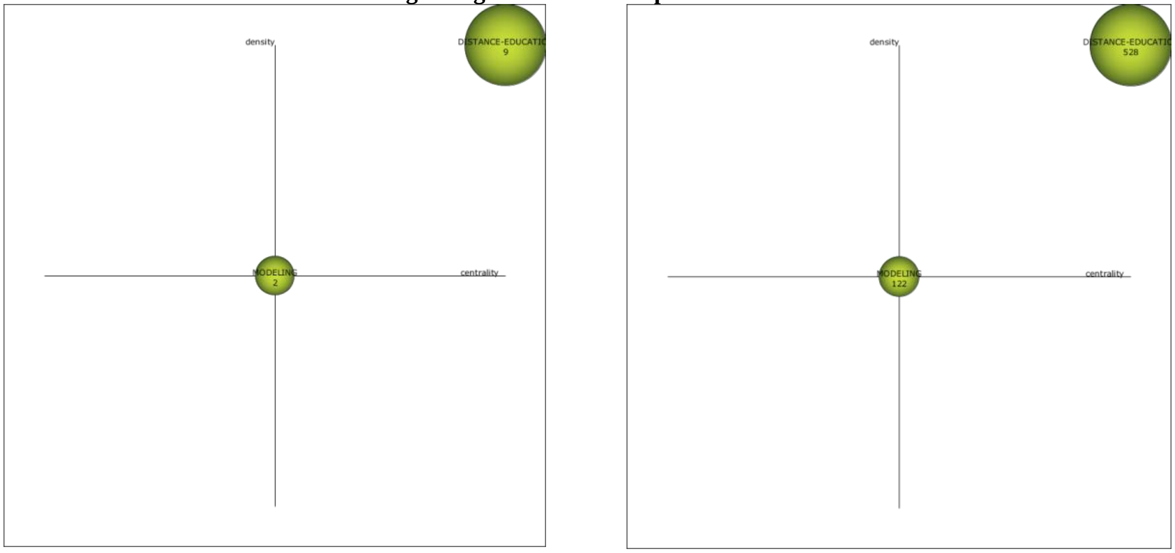
Strategic diagram based on the number of published documents

Strategic diagram based on the number of citations received

Following this pattern, in the period 2001-2005 (Figure 7 and Table 1), two themes stand out: "DISTANCE-EDUCATION" (motor theme) and "MODELING" both with a significant number of citations (528 and 122 respectively). This time it is striking the exponential increase in the number of citations that each topic gets.

Figure 7

Strategic diagrams for the subperiod 2001-2005



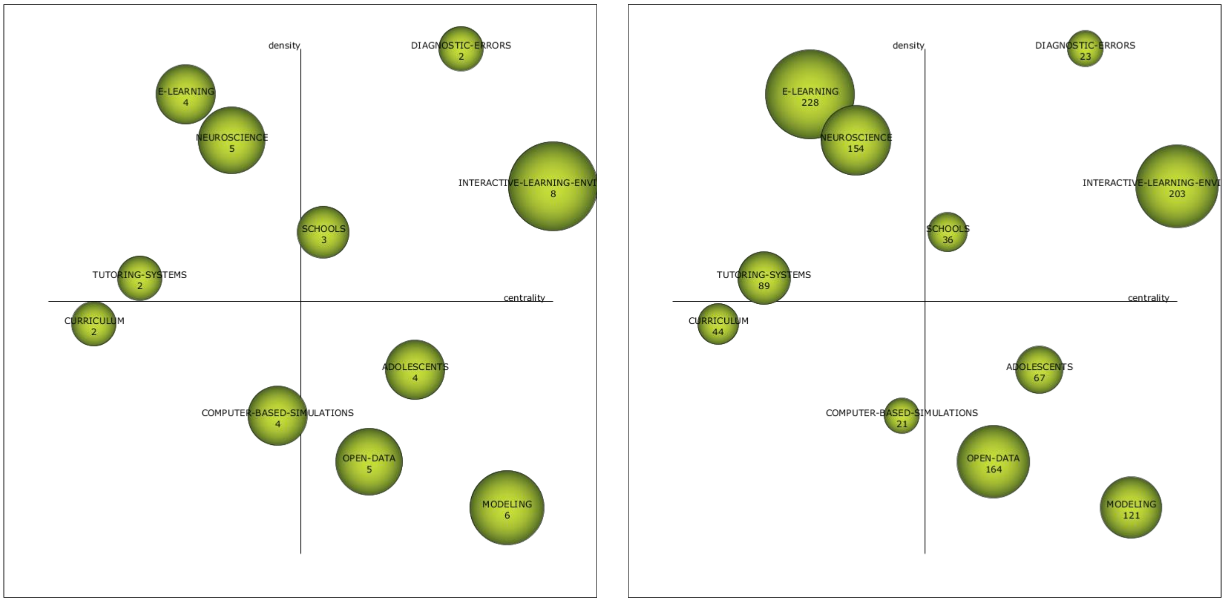
Strategic diagram based on the number of published documents

Strategic diagram based on the number of citations received

In the next period (2006-2010) (Figure 8 and Table 2) there is an explosion in thematic diversification when the field takes a decisive impulse. The main motor theme is "INTERACTIVE-LEARNING-ENVIRONMENT" with the strongest impact index. On the other hand, “MODELING” appears located in the quadrant of the basic and transversal themes together with “OPEN-DATA”. Among the highly developed but isolated themes, there are “NEUROSCIENCE” and “E-LEARNING” with great presence in the citations (especially the latter). Likewise, the presence of “TUTORING-SYSTEMS” is suggestive as a non-central issue that gains depth in its internal development. Finally, among the emerging topics of the period, “COMPUTER-BASED-SIMULATIONS” and the concept of “CURRICULUM” are recorded.

Figure 8

Strategic diagrams for the subperiod 2006-2010



Strategic diagram based on the number of published documents

Strategic diagram based on the number of citations received

Table 2

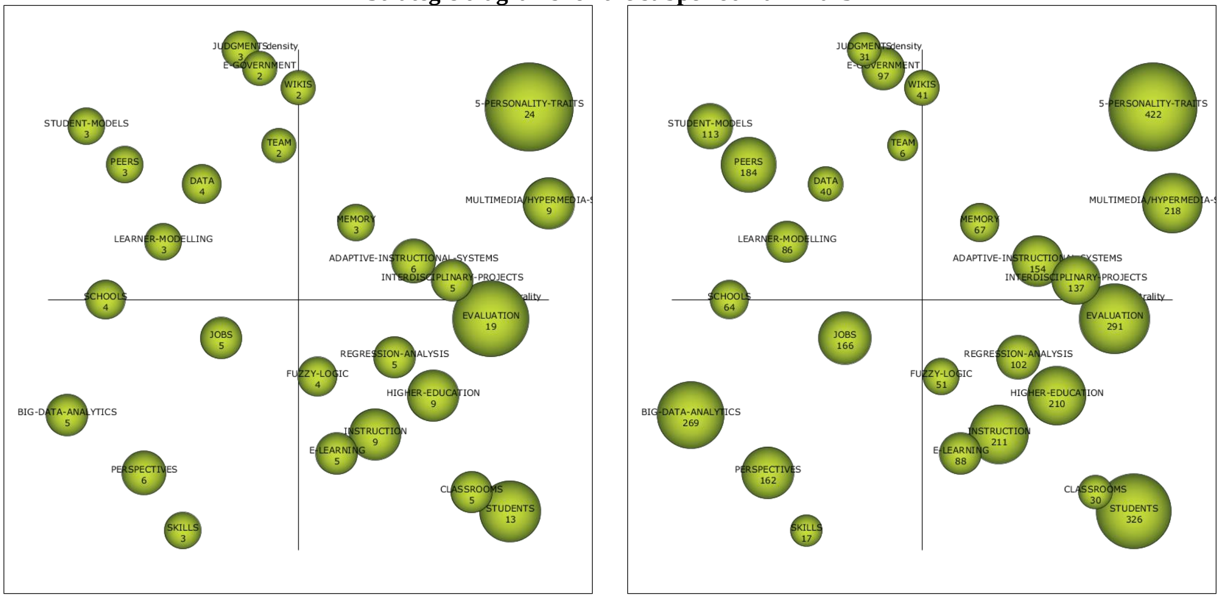
Period 2006-2010

Theme Name	Number of Documents	h-Index	Number of citations
INTERACTIVE-LEARNING-ENVIRONMENTS	8	7	203
MODELING	6	4	121
OPEN-DATA	5	5	164
NEUROSCIENCE	5	4	154
E-LEARNING	4	4	228
COMPUTER-BASED-SIMULATIONS	4	3	21
ADOLESCENTS	4	4	67
SCHOOLS	3	3	36
TUTORING-SYSTEMS	2	2	89
DIAGNOSTIC-ERRORS	2	2	23
CURRICULUM	2	1	44

In the 2011-2015 period (Figure 9 and Table 3), 26 subjects were observed, 15 more than the previous period. “5-PERSONALITY-TRAITS” stands out as an established topic with the greatest impact followed by “EVALUATION”, a topic that is entering the quadrant of the central and densely developed themes. At the same time, “MULTIMEDIA / HYPERMEDIA-SYSTEMS”, “ADAPTIVE-INSTRUCTIONAL-SYSTEMS” and “INTERDISCIPLINARY-

PROJECTS” are consolidated in central places of the research field with great citation support. “STUDENTS”, “INSTRUCTION”, and “HIGHER-EDUCATION” are located among the basic and transversal themes which receive a strong quantity of citations. It is interesting the entrance of “BIG-DATA-ANALYTICS” as an emerging topic with a considerable amount of citations, as well as “PERSPECTIVES” and “JOBS” appear in the same quadrant. Regarding highly developed but isolated topics, the presence of “STUDENT-MODELS” and “LEARNING-MODELLING” is important if we consider them together.

Figure 9
Strategic diagrams for the subperiod 2011-2015



Strategic diagram based on the number of published documents

Strategic diagram based on the number of citations received

Table 3
Period 2011-2015

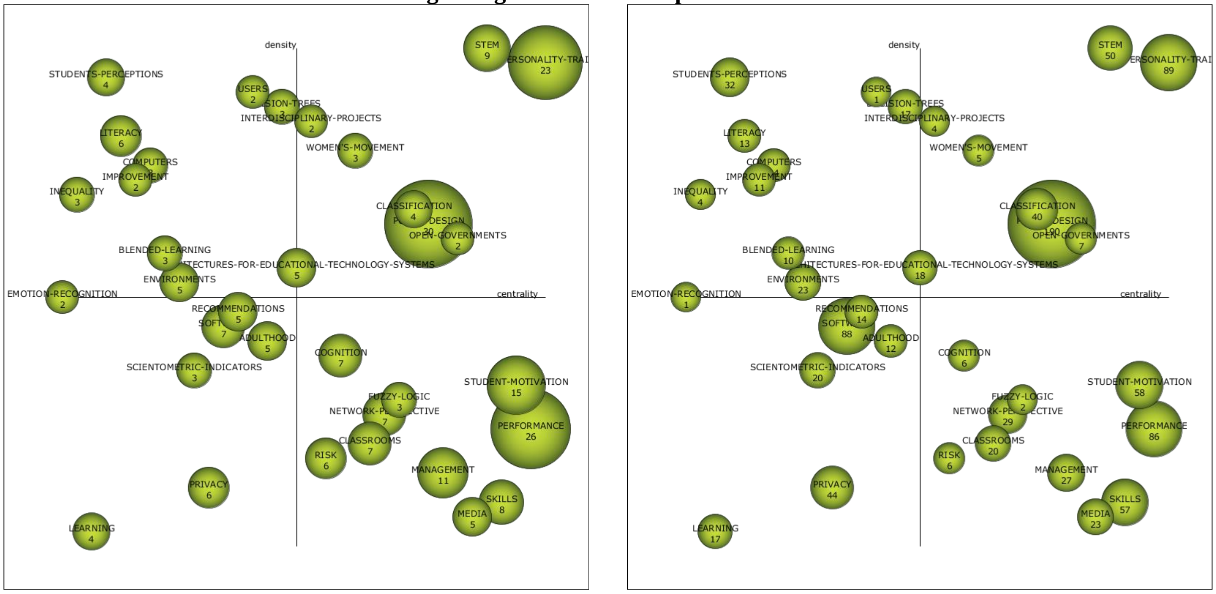
Theme Name	Number of Documents	h-Index	Number of citations
5-PERSONALITY-TRAITS	24	12	422
EVALUATION	19	10	291
STUDENTS	13	7	326
MULTIMEDIA/HYPERMEDIA-SYSTEMS	9	8	218
HIGHER-EDUCATION	9	6	210
INSTRUCTION	9	7	211
ADAPTIVE-INSTRUCTIONAL-SYSTEMS	6	5	154
PERSPECTIVES	6	4	162
INTERDISCIPLINARY-PROJECTS	5	5	137
REGRESSION-ANALYSIS	5	5	102
JOBS	5	4	166
CLASSROOMS	5	3	30
E-LEARNING	5	4	88
BIG-DATA-ANALYTICS	5	5	269
FUZZY-LOGIC	4	4	51
SCHOOLS	4	4	64
DATA	4	3	40
JUDGMENTS	3	2	31
STUDENT-MODELS	3	2	113
MEMORY	3	2	67

PEERS	3	3	184
LEARNER-MODELLING	3	3	86
SKILLS	3	3	17
E-GOVERNMENT	2	2	97
WIKIS	2	1	41
TEAM	2	1	6

The period 2016-2019 (Figure 10 and Table 4) has 34 themes showing sustained growth. The thematic proliferation is undeniable. “5-PERSONALITY-TRAITS” motor theme is maintained but now “POLICY-DESIGN” is gaining presence with a large number of references. Among the most prominent basic themes are “PERFORMANCE”, “STUDENT-MOTIVATION” and “MANAGEMENT”, as well as “PRIVACY” appears as an interesting emerging topic in the period. The highly developed themes but little connected with others are multiple but few documents refer to them and they have only a few citations (with exception of “STUDENT-PERCEPTIONS” and “ENVIRONMENTS” which have more references).

Figure 10

Strategic diagrams for the subperiod 2016-2019



Strategic diagram based on the number of published documents

Strategic diagram based on the number of citations received

Table 4

Period 2016-2019

Theme Name	Number of Documents	h-Index	Number of citations
POLICY-DESIGN	30	7	190
PERFORMANCE	26	5	86
5-PERSONALITY-TRAITS	23	6	89
STUDENT-MOTIVATION	15	4	58
MANAGEMENT	11	3	27
STEM	9	4	50
SKILLS	8	4	57
SOFTWARE	7	3	88
NETWORK-PERSPECTIVE	7	3	29
COGNITION	7	1	6

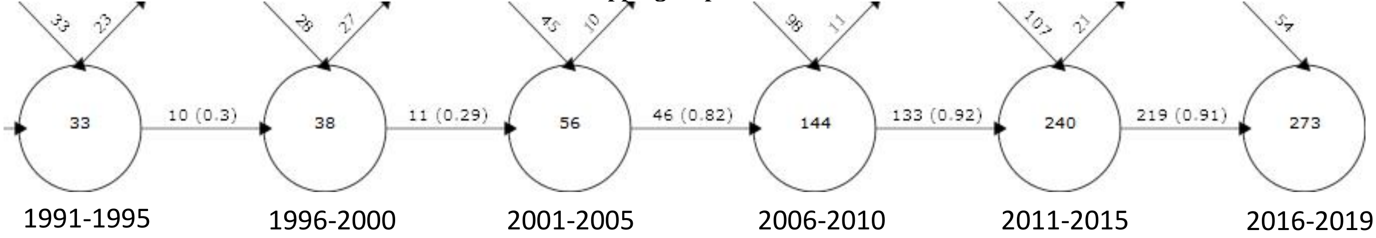
CLASSROOMS	7	3	20
LITERACY	6	1	13
RISK	6	1	6
PRIVACY	6	3	44
ADULTHOOD	5	2	12
ARCHITECTURES-FOR-EDUCATIONAL-TECHNOLOGY-SYSTEMS	5	2	18
RECOMMENDATIONS	5	1	14
MEDIA	5	2	23
ENVIRONMENTS	5	3	23
CLASSIFICATION	4	3	40
STUDENTS-PERCEPTIONS	4	2	32
LEARNING	4	2	17
WOMEN'S-MOVEMENT	3	1	5
COMPUTERS	3	2	14
DECISION-TREES	3	2	17
BLENDED-LEARNING	3	2	10
INEQUALITY	3	1	4
FUZZY-LOGIC	3	1	2
SCIENTOMETRIC-INDICATORS	3	3	20
INTERDISCIPLINARY-PROJECTS	2	2	4
IMPROVEMENT	2	2	11
USERS	2	1	1
OPEN-GOVERNMENTS	2	2	7
EMOTION-RECOGNITION	2	1	1

Stability Diagram: keywords evolution

In order to show the keywords' evolution, Figure 11 allows to represent its number for each period (number in the circle) and the number of keywords that are maintained (horizontal arrow) as well as the number of keywords that goes out (up-arrow) and those which are new incorporations (down-arrow). Also, the number in parenthesis is the Similarity Index.

In general terms, it is possible to observe clearly that the quantity of keywords increases from 33 in the period 1991-1995 until 273 in the period 2016-2019. Even it is clear the jump registered in 2006-2010 (from 56 keywords from the previous period to 144). The Similarity Index also increases drastically from 0.3 to 0.91 which means that the terminology is shared and maintained while the research field is consolidated especially in the last periods.

Figure 11
Overlapping map



Thematic Evolution

After analyzing the evolution of the keywords, it is necessary to consider the thematic evolution. As mentioned previously, each theme is composed of a particular network of keywords.

Doing so, regarding the repetition of themes between periods (see Table 5), it is noted that mostly there are continuous themes between 2 consecutive time periods but no more than this. Only in 9 opportunities, thematic continuities are detected between periods. Particularly, 5 repeated themes are recorded in the 2011-2015 and 2016-2019 periods, where not only the thematic proliferation but also the maintenance of interest on specific issues begin to be noticed. The latter gives us signs about the beginning of the research field strengthening, although it must be said that, in general, in recent periods many themes appear but few consolidated by continuity.

Table 5
Repeated Themes

Theme Name	Number of Documents	h-Index	Number of citations	Repeated period
COMPUTER-BASED-SIMULATIONS	2	2	27	1996-2000 & 2006-2010
MODELING	2	2	122	2001-2005 & 2006-2010
E-LEARNING	4	4	228	2006-2010 & 2011-2015
SCHOOLS	3	3	36	2006-2010 & 2011-2015
5-PERSONALITY-TRAITS	24	12	422	2011-2015 & 2016-2019
INTERDISCIPLINARY-PROJECTS	5	5	137	2011-2015 & 2016-2019
FUZZY-LOGIC	4	4	51	2011-2015 & 2016-2019
CLASSROOMS	5	3	30	2011-2015 & 2016-2019
SKILLS	3	3	17	2011-2015 & 2016-2019

According to the Thematic Evolution Diagram (Figure 12), a high density of links stands out especially between the last 3 periods 2006-2010, 2011-2015 and 2016-2019.

Dotted lines predominate, which means that the linked topics share keywords that are not the name of the theme (the central keyword in the network). This suggests many thematic cross-links (especially from 2006 onwards) which prevents talking about many consolidated lines of interest for research in the field. In fact, this means that the topics feed each other, showing little cohesion and, therefore, little definition of clear thematic lines. For example, a representative case of this situation is the theme "PERFORMANCE", which is very important in 2016-2019 in quantitative terms, but that nurtures from several previous topics ("STUDENTS", "LEARNING-MODELING", etc.) that occupy a peripheral position in its constitution as a theme.

However, there are some continuous thematic areas expressed by solid lines in which the linked themes share a name (this includes both themes that show the same name or the name of one is part of the other theme). For example, the line "NEUROSCIENCE" > "5-PERSONALITY-TRAITS" > "5-PERSONALITY-TRAITS" is a strong thematic line with presence from 2006 onwards not only because of the similarity of its thematic composition but also because of the considerable number of documents that refer to it over time. Similarly, the

"DATA" > "POLICY-DESIGN" line shows strong links since 2011 and an increase in the last node comparing to the previous one.

It is interesting the strong overlapping between "MODELING" (2006-2010) with the theme "EVALUATION" (2011-2015) that stands out for the number of documents that refer to this latter topic. On the other hand, it is also highlighted the link between "REGRESSION-ANALYSIS" in the period 2011-2015 with "CLASSIFICATION" in the period 2016-2019 (with a considerable Inclusion Index although not many documents refer to these thematic areas in those periods).

Table 6
Main Themes per period
(by number of documents associate to them)

Period	Theme Name	Number of Documents	h-Index	Number of citations
1991-1995	ARTIFICIAL-INTELLIGENCE	5	2	17
1996-2000	LEARNING-ENVIRONMENTS	3	3	21
2001-2005	DISTANCE-EDUCATION	9	8	528
2006-2010	INTERACTIVE-LEARNING-ENVIRONMENTS	8	7	203
2011-2015	5-PERSONALITY-TRAITS	24	12	422
2016-2019	POLICY-DESIGN	30	7	190

In Table 6 is possible to observe the main thematic areas for each period from 1991 onwards. First, the focus is on artificial intelligence systems in education under the domain of Computer Sciences topics. Second, a more pedagogical concern is predominant with interest in learning and distance education. Finally, it is recorded a strong presence of thematic areas related to the intelligent detection of personality traits as well as, more recently, themes of policy design using large amounts of data.

Having said that, if the composition of these main themes is analyzed in greater depth, it is possible to consider the network of keywords that integrate them.

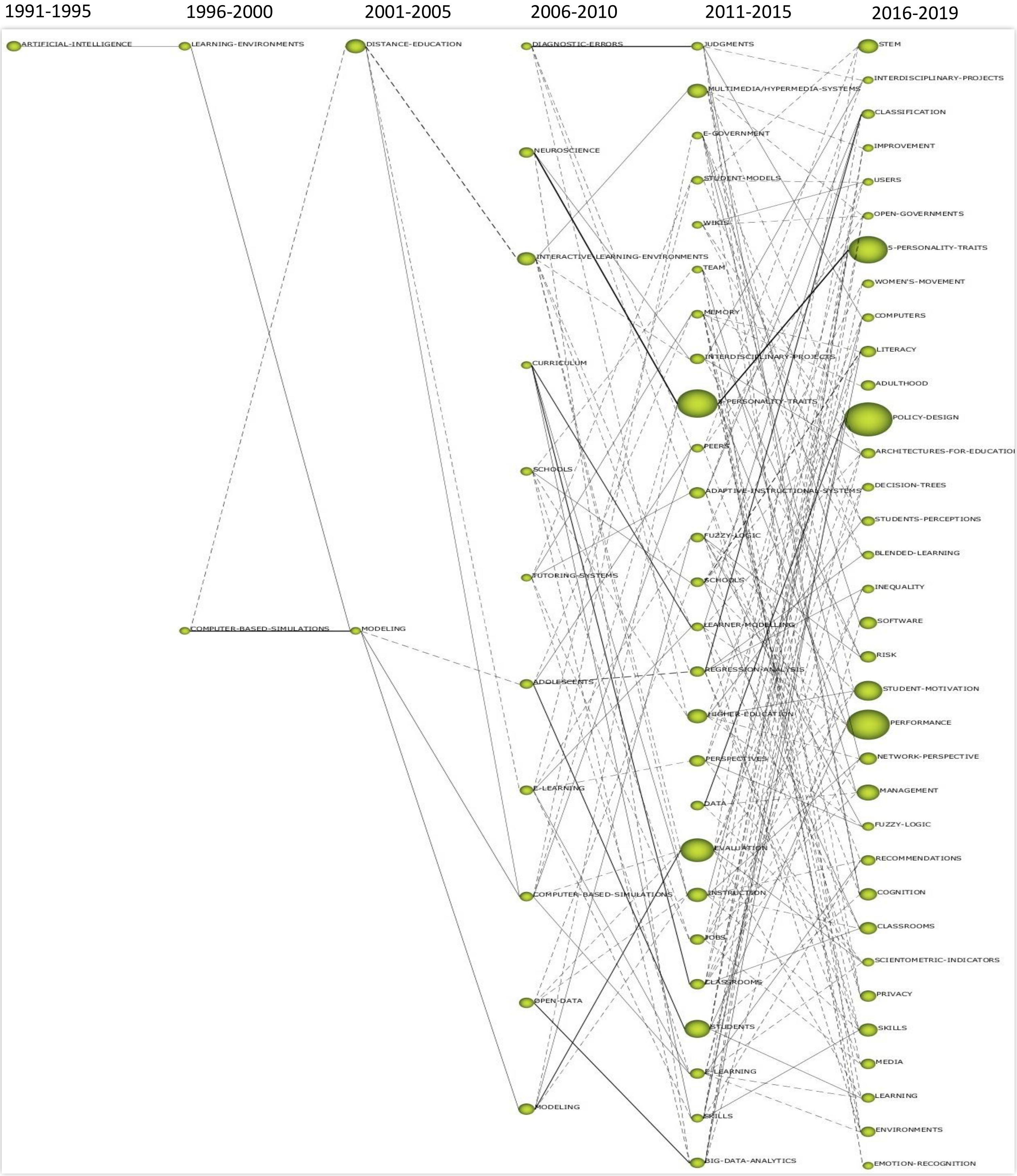
In the case of the 1991-1995 period (Figure 13, left), the theme "ARTIFICIAL INTELLIGENCE" has, naturally, the keyword of the same name in the center of the network and it accumulates the greatest amount of mentions in the analyzed documents. In addition, the strongest link in this network is with "Expert Systems".

Briefly, it is observed that in this period the centrality of a keyword from the Computer Science field stands out and, as peripheral, some keywords typically identified with the production of the pedagogical field such as "Evaluation".

In the following period (1996-2000) (Figure 13, right), the keyword "Artificial Intelligence" is the most referred to but not the central one anymore. Here, the centrality is occupied by the keyword "Learning Environments" that has a strong link (high Equivalence Index) with "Artificial Intelligence".

Figure 12

Thematic Evolution Diagram



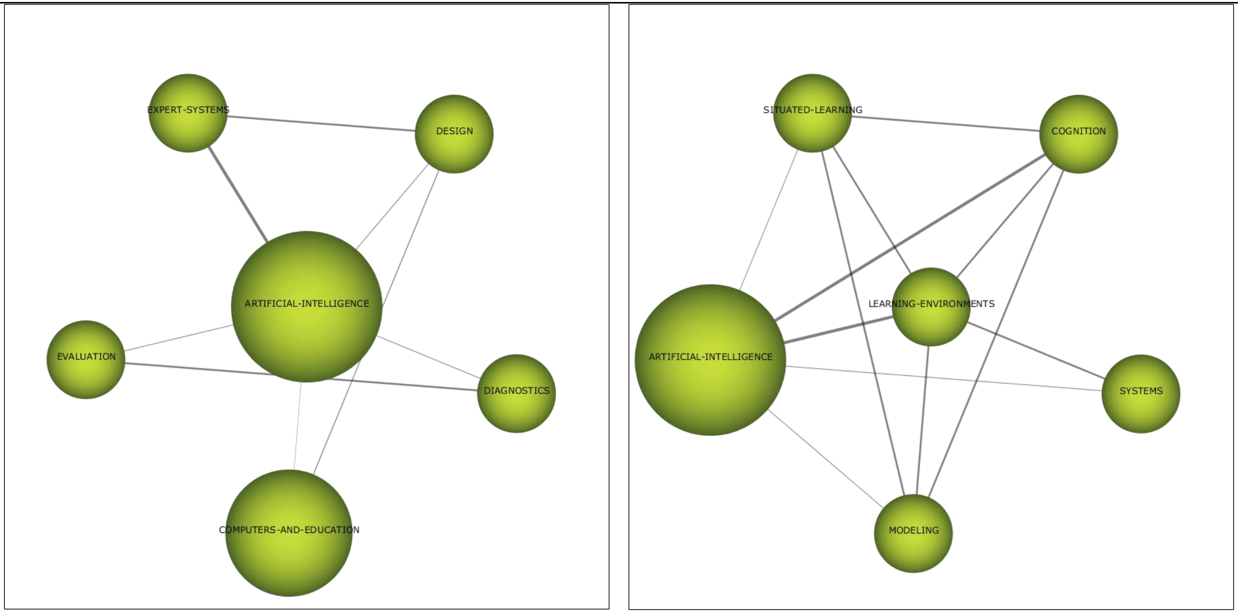
During this time a significant variation is noticed: the category “Artificial Intelligence” goes to the periphery and “Learning” is located in the center. This is clearly suggestive of the intervention of new logics that enter the field of scientific production here studied.

Figure 13

Keyword configuration per each main theme per period

(1991-1995)
ARTIFICIAL-INTELLIGENCE

(1996-2000)
LEARNING-ENVIRONMENTS



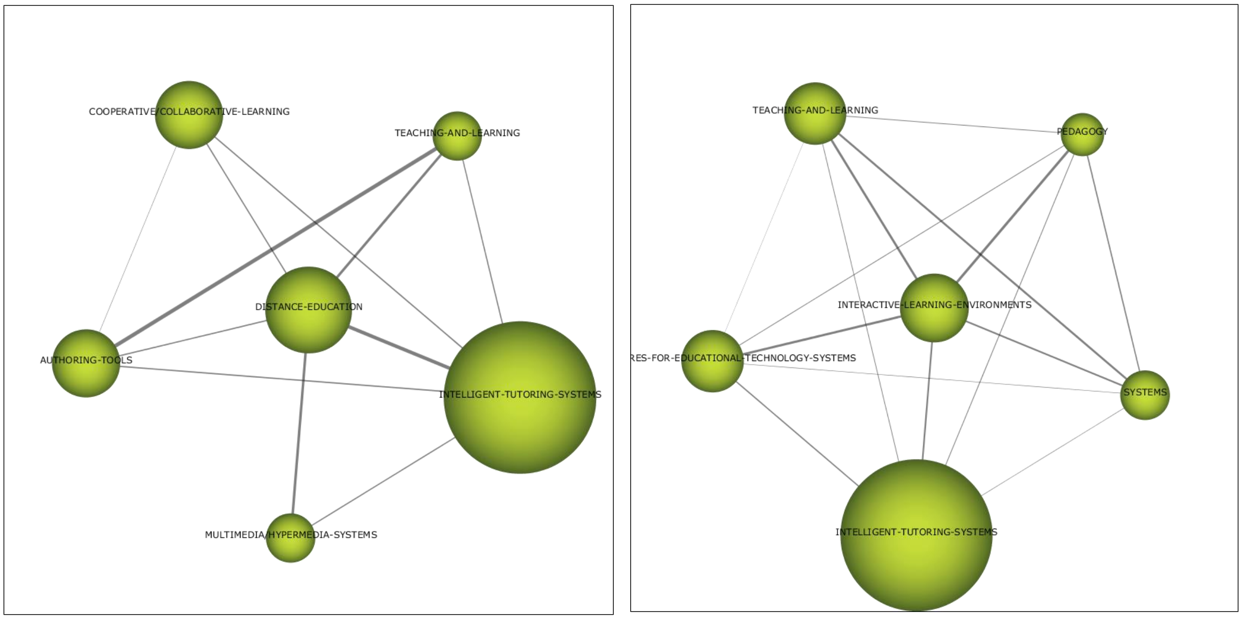
In the period 2001-2005 (Figure 14, left), “Distance Education” occupies the central place but the most referred keyword in the documents of the period is “Intelligent Tutoring Systems”. Between these two nodes, there is a considerable Equivalence Index expressed in the thickness of the line. At the same time, in this thematic area, the link between “Authoring Tools” and “Teaching and Learning” stands out.

In the period 2006-2010 (Figure 14, right), whose most important theme is “INTERACTIVE-LEARNING-ENVIRONMENTS”, the keyword of the same name is central but not the most mentioned in the documents. The most mentioned keyword is, as in the previous period, “Intelligent Tutoring Systems” and the strongest links of the central node are given with typically pedagogical categories: “Teaching and Learning”, “Pedagogy”, “Resources for Education Technology Systems”.

In these last two periods, the pedagogical keywords are central, while some of the peripheral keywords come from the field of Computer Sciences (such is the case of the “Intelligent Tutoring Systems” which anyway has the most important number of mentions in both periods).

Figure 14

Keyword configuration per each main theme per period


COOPERATIVE/COLLABORATIVE-LEARNING
TEACHING-AND-LEARNING
DISTANCE-EDUCATION
AUTHORING-TOOLS
INTELLIGENT-TUTORING-SYSTEMS
MULTIMEDIA/HYPERMEDIA-SYSTEMS
TEACHING-AND-LEARNING
PEDAGOGY
INTERACTIVE-LEARNING-ENVIRONMENTS
RES-FOR-EDUCATIONAL-TECHNOLOGY-SYSTEMS
SYSTEMS
INTELLIGENT-TUTORING-SYSTEMS

(2001-2005)
DISTANCE-EDUCATION

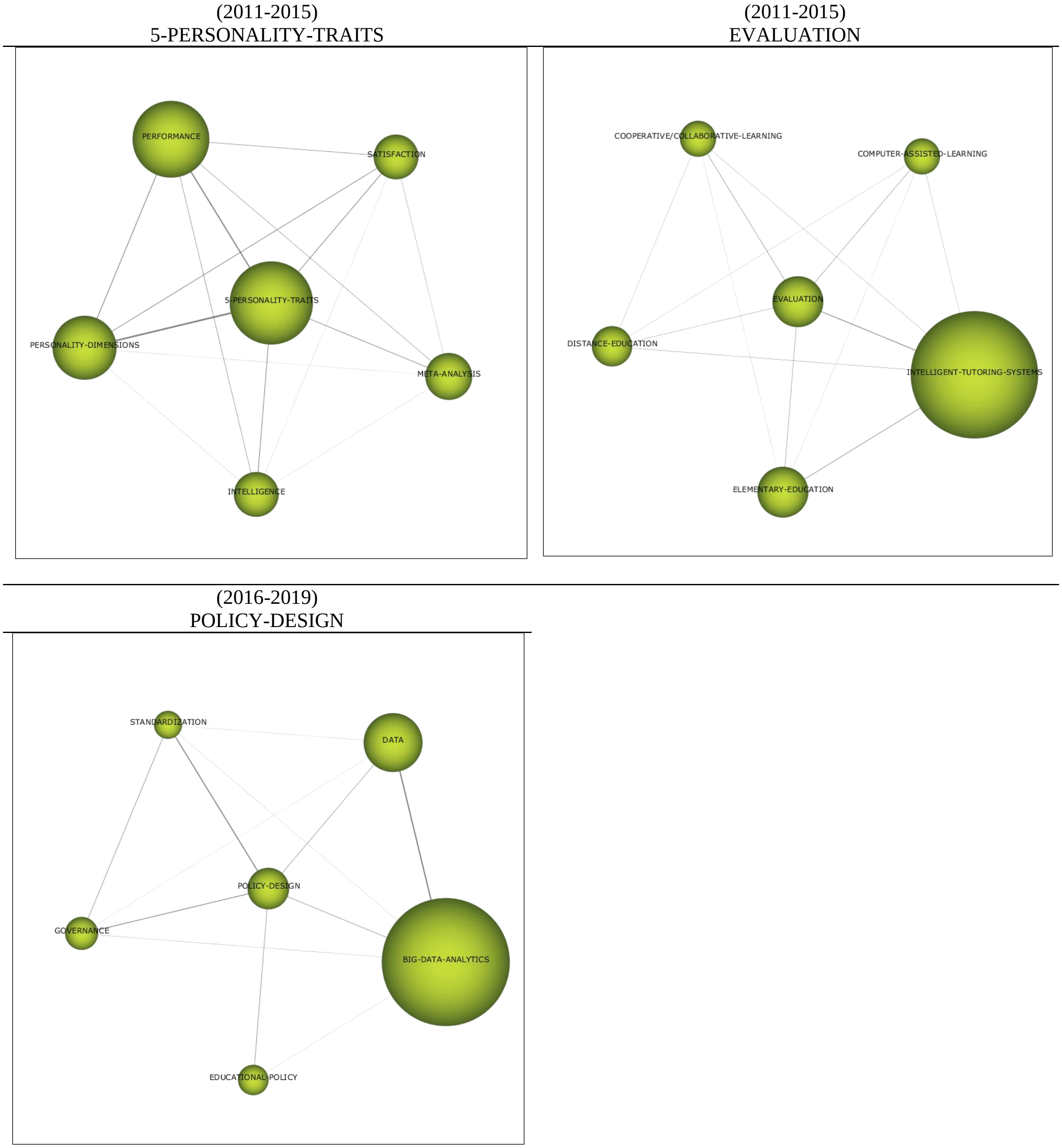
(2006-2010)
INTERACTIVE-LEARNING-ENVIRONMENTS

In the 2011-2015 period (Figure 15, left), the main theme is "5-PERSONALITY-TRAITS" and the configuration of its thematic network represents meta-analytical works aimed at the intelligent detection of personality traits according to the students' performance. In addition, it is important to highlight that, in this period (Figure 15, right), the theme "EVALUATION" has considerable mentions in the documents and its conformation also shows the significant presence of the keyword "Intelligent Tutoring Systems".

Finally, in the 2016-2019 period (Figure 15, below), the topic "POLICY-DESIGN" is composed of a keywords network where "Big Data Analytics" and "Data" are important categories related to those educational policy design.

Figure 15

Keyword configuration per each main theme per period



(2011-2015)
5-PERSONALITY-TRAITS

(2011-2015)
EVALUATION

COOPERATIVE/COLLABORATIVE-LEARNING
COMPUTER-ASSISTED-LEARNING
EVALUATION
DISTANCE-EDUCATION
INTELLIGENT-TUTORING-SYSTEMS
ELEMENTARY-EDUCATION

(2016-2019)
POLICY-DESIGN

Conclusions

This exploratory study performed a Systematic Literature Review (SLR) using text mining technics applied over bibliographic material with SciMAT software. The objective was double: a) to specify what are the main topics of Artificial Intelligence investigated in the field of educational research during the last three decades; b) to describe which is its thematic evolution during this period.

Among the main obtained results, it can be mentioned that: (1) academic production about Artificial Intelligence in the educational field prevails in the last decade, with sustained growth; (2) the leading countries that take part in this type of scientific production are the United States (predominantly), United Kingdom, China, and Spain; (3) it is observed a drastic thematic proliferation (especially from 2006 onwards) with incipient consolidation of a few solid lines of investigation referred to automatic detection of personality traits and educational policy design using big data; (4) two thematic circuits are distinguished (one more technical with keywords coming from Computer Sciences and the other one with pedagogical/psychological concerns). These big conceptual organizers show different emphasis and maintain different crossings during the analyzed periods: first, the focus is on Computer Sciences topics (1991-1995); second, a more pedagogical concern is established in the field (1996-2010) with strong but peripheric presence of Intelligent Tutoring Systems developing; third, a big production around intelligent systems to detect personality features is developed (2011-2015); forth, the currently interest is over policy design using large amounts of data (2016-2019).

For future studies, it is considered necessary to deepen on other specific themes composition like “Classification”, “Privacy”, and “Risk” according to some historic-line of landmarks that would allow enriching the analysis from a qualitative point of view. Also, for confirmatory purposes, further analysis could be done based on other bigger databases that are already available and downloaded. In addition, the comparison with Latin American academic production would be suggestive.

Bibliography

A. Breiter, “Datafication in education: a multi-level challenge for IT in educational management,” in Stakeholders and Information Technology in Education, T. Brinda, N. Mavengere, I. Haukijarvi, C. Lewin, D. Passey Eds. Berlin: Springer, 2016, pp. 95-103.

B. K. Daniel, Ed., Big data and Learning Analytics in higher education: current theory and practice. Switzerland: Springer International Publishing, 2017.

B. Williamson, Big data in education: the digital future of learning, policy and practice. London: SAGE, 2017.

C. Okoli and K. Schabram, “A guide to conducting a systematic literature review of information systems research”, Sprouts: Working Papers on Information Systems, vol. 10, num. 26, pp. 1-51, 2010.

D. Kehl, P. Guo and S. Kessler, Algorithms in the criminal justice system: assessing the use of risk assessments in sentencing. Berkman Klein Center for Internet & Society, Harvard Law School, 2017.

E. Siegel, Predictive Analytics. The power to predict who will click, buy, lie or die. New Jersey: John Wiley and Sons, 2016.

G-H. Kim, S. Trimi, J-H. Chung, “Big-Data Applications in the Government Sector,” Communications of the ACM, vol. 57, num. 3, pp. 78-85, 2014.

J. Danaher, M. J. Hogan, C. Noone, R. Kennedy, A. Behan, A. De Paor, H. Felzmann, M. Haklay, S. Khoo, J. Morison, M. H. Murphy, N. O'Brolchain, B. Schafer and K. Shankar, “Algorithmic governance: Developing a research agenda through the power of collective intelligence,” Big Data & Society, pp. 1–21, 2017.

J. Meca, “Cómo realizar una revisión sistemática y un meta-análisis,” Aula abierta, vol. 38, num. 2, pp.53-64, 2010.

J. Van Dijck, “Datafication, dataism and dataveillance: big data between scientific paradigm and ideology,” Surveillance & Society, vol. 12, num. 2, pp. 197-208, 2014.

K. Jee, G-H. Kim, “Potentiality of big data in the medical sector: focus on how to reshape the healthcare system,” Healthc Inform Res, vol. 19, num. 2, pp. 79-85, 2013.

M. Batty, “Big data, smart cities and city planning,” Dialogues in Human Geography, vol. 3, num. 3, pp. 274–279, 2013.

M.A Martínez, M.J. Cobo, M. Herrera and E. Herrera-Viedma, “Analyzing the Scientific Evolution of Social Work Using Science Mapping”, Research on Social Work Practice, vol. 25, num. 2, pp. 257-277, 2015.

M.J. Cobo, A.G. López-Herrera, E. Herrera-Viedma and F. Herrera, “An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field”, Journal of Informetrics, vol. 5, num. 1, pp. 146-166, 2011.

M.J. Cobo, A.G. López-Herrera, E. Herrera-Viedma and F. Herrera, “SciMAT: A new Science Mapping Analysis Software Tool”, Journal of the American Society for Information Science and Technology, vol. 63, num. 8, pp. 1609-1630, 2012.

N. Sclater, A. Peasgood and J. Mullan, Learning Analytics in Higher Education. A review of UK and international practice. United Kingdom: Jisc, 2016.

N. Selwyn, “Data entry: towards the critical study of digital data and education,” Learning, Media and Technology, vol. 40, num. 1, pp. 64–82, 2015.

O. Simpson, “Predicting student success in open distance learning,” Open Learning, vol. 21, num. 2, pp. 125-138, 2006.

R. S. Baker, “Stupid tutoring systems, intelligent humans,” International Journal of Artificial Intelligence in Education, vol. 26, Issue 2, pp. 600-614, 2016.

V. Aleven, J. Sewall, O. Popescu, F. Xhakaj, D. Chand, R. Baker, Y. Wang, G. Siemens, C. Rosé and D. Gasevic, “The Beginning of a Beautiful Friendship? Intelligent Tutoring Systems and MOOCs” in Artificial Intelligence in Education, AIED 2015, C. Conati, N. Heffernan, A. Mitrovic, M. Verdejo, Eds. Lecture Notes in Computer Science, vol 9112. Springer, 2015, pp. pp 525-528.

W. Holmes, M. Bialik, and C. Fadel, Artificial intelligence in education: promises and implications for teaching and learning. Boston, MA: The Center for Curriculum Redesign, 2019.

SciMAT webpage: https://sci2s.ugr.es/scimat/index.html