

# Systematic review

## “Human Capital value in Hi-Tech Industry”

Group H

Stefania Sferragatta  
Eugenio Baldo  
Clara Lecce

April 2021

## 1 Introduction

The aim of our systematic review is to identify, appraise and synthesise all relevant studies on the following topic: *Human Capital value in technologies firms*.

We can divide this work in stages, indeed firstly we specified the objectives of the review, then we found the relevant research and collected the data, after we assess quality of studies and in the end we synthesise evidence in order to interpret findings and to draw the conclusions.

In the following sections we are going to explain in detail the steps we made in order to obtain our results.

## 2 Eligibility criteria

Our first step is the specification of criteria for including and excluding studies in the review related to the chosen topic.(1)

The criteria consist of: select only sources in English language whose area of interest is related to ‘science and technology’ and expand the research also to related words and equivalent topics.

## 3 Information sources

The main information sources used are: Web of Science, Scopus and SBS EB-SCO.

They are citation and abstraction's databases that provide comprehensive coverage of the peer-reviewed journal and conference literature, with links to full-text (where available) through the libraries.

## 4 Model for conceptualizing the queries

To draw our qualitative study we used the SPICE model (2), which allows us to answer the following research question:

**Which are the most related publications to the research domain  
“Human intellectual capital in tech industry”?**

- Setting - Where?  
Scopus, WOS, SBS Ebsco.
- Perspective - For whom?  
Who is interested in Human Capital in the digital field.
- Intervention - What?  
Books, articles, papers and so on related to the topic.
- Comparison - What else?  
None systematic review related to this topic.
- Evaluation - How well?  
Through a score function based on text processing.

We used the Boolean operators (AND, OR) to combine concepts and the asterisk symbol \* at the stem of a word to provide spelling variations.(3)

The approach we used is mainly based on **keyword searching** in order to obtain records that contain a given word or phrase in specified fields of the database record.(4) This allows us to speed up research and to decrease the number of documents retrieved, but to restrict the domain of interest we had to use also other strategies that will be explained in detail in the following section.

## 5 Search strategy

In this section we list the strategies applied in the Information sources and we show the queries done. Even if the queries used on the sources are different, since the structure also depends on the Information source, they all have in common some principles:

- The topics searched are: 'Human Capital', 'technology', 'firms' and all the combinations of them;
- The subject area chosen are: 'Computer science', 'Business', 'Social/Decision science'.

In performing the search on SBS EBSCO we also applied other limitations due to the huge amount of documents retrieved:

- The publication date must be between the 1990 and 2021;
- Specify the documents' sources.

In the following list there are the queries used:

- **Scopus Query 1:**

```
(TITLE-ABS-KEY ( human AND capital ) AND TITLE-ABS-KEY ( technolog*)  
AND TITLE-ABS-KEY ( firm* ) OR TITLE-ABS-KEY ( compan* ) )  
AND ( LIMIT-TO ( EXACTKEYWORD , "Human Capital" ) OR LIMIT-TO  
( EXACTKEYWORD , "Intellectual Capital" ) OR LIMIT-TO ( EXACTKEYWORD  
, "Knowledge Management" ) OR LIMIT-TO ( EXACTKEYWORD , "Human  
Capitals" ) OR LIMIT-TO ( EXACTKEYWORD , "Information Technology")  
OR LIMIT-TO ( EXACTKEYWORD , "Human Resource Management" )  
OR LIMIT-TO ( EXACTKEYWORD , "Industry" ) OR LIMIT-TO ( EXACTKEYWORD  
, "Technology" ) OR LIMIT-TO ( EXACTKEYWORD , "Technology Transfer"  
) OR LIMIT-TO ( EXACTKEYWORD , "Productivity") OR LIMIT-TO  
( EXACTKEYWORD , "Social Capital" ) OR LIMIT-TO ( EXACTKEYWORD  
, "Technological Development" ) OR LIMIT-TO ( EXACTKEYWORD ,  
"Personnel Training") OR LIMIT-TO ( EXACTKEYWORD , "R&D" )  
OR LIMIT-TO ( EXACTKEYWORD , "Technological Innovation" ) OR  
LIMIT-TO ( EXACTKEYWORD , "Economic Growth" ) OR LIMIT-TO  
( EXACTKEYWORD , "Firm Performance" ) OR LIMIT-TO ( EXACTKEYWORD  
, "Social Capitals" ) OR LIMIT-TO ( EXACTKEYWORD , "Skilled  
Labor" ) OR LIMIT-TO ( EXACTKEYWORD , "Technological Change"  
) OR LIMIT-TO ( EXACTKEYWORD , "Relational Capital" ) OR LIMIT-TO  
( EXACTKEYWORD , "Empirical Analysis" ) OR LIMIT-TO ( EXACTKEYWORD  
, "Industry 4.0" ) OR LIMIT-TO ( EXACTKEYWORD , "Intellectual  
Property" ) OR LIMIT-TO ( EXACTKEYWORD , "Cost Benefit Analysis"  
) OR LIMIT-TO ( EXACTKEYWORD , "Labor Productivity" ) OR LIMIT-TO  
( EXACTKEYWORD , "Big Data" ) OR LIMIT-TO ( EXACTKEYWORD ,  
"Digital Transformation" ) OR LIMIT-TO ( EXACTKEYWORD , "Skills"  
)) AND ( LIMIT-TO ( SUBJAREA , "COMP" ) ) AND ( LIMIT-TO  
SUBJAREA , "BUSI" ) OR LIMIT-TO ( SUBJAREA , "SOCI" ) OR  
LIMIT-TO ( SUBJAREA , "DECI" ) ) AND ( LIMIT-TO ( EXACTSRCTITLE  
, "International Journal Of Technology Management" ) OR LIMIT-TO  
( EXACTSRCTITLE , "Proceedings Of The International Conference  
On Intellectual Capital Knowledge Management And Organisational  
Learning Icickm" ) )
```

- **Web Of Science**

**Query 1 from all database:**

(ORDER BY ‘TIMES CITED’ ): TOPIC: (Human Capital) AND TOPIC: (technolog\*) AND TOPIC: (firm\*) Refined by: RESEARCH DOMAINS: ( SCIENCE TECHNOLOGY ) Timespan: All years. Databases: WOS, CCC, DRCI, DIIDW, KJD, MEDLINE, RSCI, SCIELO, ZOOREC. Search language=EN

#### Query 2:

TOPIC: (Human Capital) AND TOPIC: (technolog\*) AND TOPIC: (firm\*) Refined by: RESEARCH DOMAINS: ( SCIENCE TECHNOLOGY ) Timespan: All years. Databases: WOS. Search language=EN

- SBS EBSCO

#### Query 1:

Find any term of the search:

TX human capital AND TX technologies firms AND TX Human Capital value in technologies firms

Expanders:

Apply Related words, Apply equivalent topics

Limitors:

Publication Date: 1990-2021

Arguments:

Human capital, Economic development, Personnel management, Technological innovations,

technology, information technology, intellectual capital, social capital, human resource, management, research & development, hr & organizational behaviours, human capital - management,

Language: English

Pubblication:

International Journal of Human Resource Management, personnel review, Journal of Intellectual Capital, r&d management, Technology Analysis & Strategic Management.

## 6 Data Visualization

For both visualizations we used the VOSviewer software.

The type of analysis chosen is co-occurrence, hence the relatedness of items is determined based on the number of documents in which they occur together. (5) The counting method used is full and the minimum number of occurrences of a keyword has been selected equal to 5 (thresholds) thereby the total strength of the co-occurrence links with other keywords is computed for each of the keywords that passed that threshold. The weights applied is based on the *total link strength* and the method analysis is the *Association strength* one.

For both networks, the cluster with more connected links is related to the keyword *Human capital* and it is a reliable result and in agreement with our research.

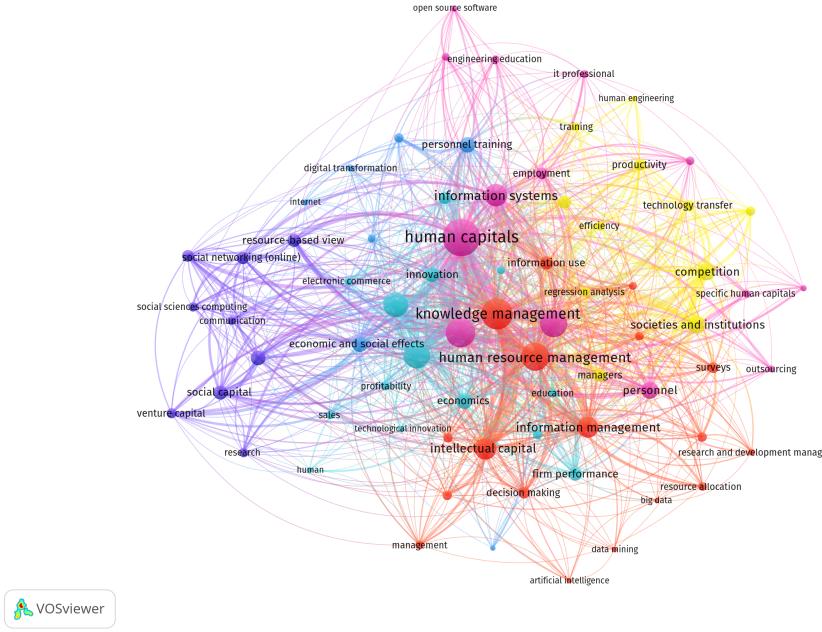


Figure 1: Scopus network

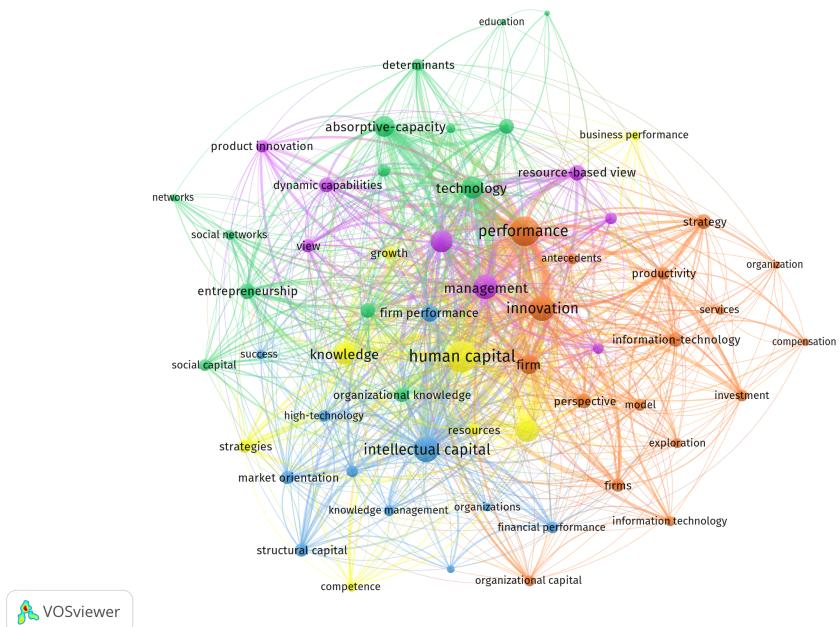


Figure 2: WOS network

## 7 Merging

For each Database, we used a specific query to obtain results as described above. In the following list we provide the number of results for each query:

- Scopus database - 203 publications
- EBSCO database - 34 publications
- WOS database - 154 publications

In order to have a comprehensive view of all the results, we have merged the results of the different databases using Python<sup>1</sup>. The merge of the results has been performed using Pandas library(6).

For the merged dataset, we decide to use the following fields.

- Authors - a list of the publication's authors
- Title - the title of the publication
- Year - the year when the publication was released
- References - the other publications cited by the paper
- Abstract - a synopsis of the publication
- Subject - the main research area of the publication
- Database - the source database of the publication

However, not all the attributes were included in each database.

For the Scopus database, there was not provided the *subject* field; for the EBSCO database there were not provided the *reference* field and the *subject* field and for the WOS database there was not provided the *reference* field.

When the field *reference* was absent, it was replaced by *no reference*, as well as for the *subject* field it was replaced by *no subject*. The merge has been accomplished by concatenating vertically the processed datasets.

Afterwards, duplicates have been dropped using a function provided in the Pandas Package. In this way, the total size of the dataset has been reduced from 391 publications to 373 distinct ones.

## 8 Selection process

Our goal is to select the most pertinent publications to our research domain using a text processing approach. We focused on the abstract of each publication.

Being the abstract a short description of the publication, it is very likely to find the most relevant words regarding the topic. To do our analysis we used

---

<sup>1</sup>libraries used for merging and selection steps are Pandas, NLTK ,Matplotlib, WordCloud

Python(9) and the NLTK(7) library. In order to take into account the different speech tags, we used a process called stemming, which remove the desinence of the word. Stemming was applied to each word.

Our text processing approach can be divided into the following steps:

1. Choose the most relevant words according to the research domain;
2. Stem the chosen words (we have used the PorterStemmer of NLTK). These words are the topic stems. In addition, for each stem it is assigned a weight. this weight is an integer in the range [1,3];
3. Tokenize the abstract, remove stopwords and punctuation, stem the Abstract content with the same stemmer;
4. Define a score metric to rank publications. In our case, we defined the following score frequency metric applied to each abstract:

$$\frac{\sum_{i=1}^n w_i(stem)}{3n} \quad if \quad stem \quad in \quad topic \quad stems \quad (1)$$

where n is the number of stems in the abstract and 3 the maximum weight this score takes into account both the frequency and the document length;

5. Rank the publications according to the score in ascending order;
6. Choose the top-k publications according to the score.

As a result of the above algorithm, we obtained a ranking of the publications. In the table below, we have inserted the top 9.

| Title  | Score  |
|--|--------|
| Human resource flow within software industry: a firm-level investigation   | 0.2087 |
| Complementary IT resources for enabling technological opportunism  | 0.1528 |
| Firm-specificity and organizational learning-related scale on investment in internal human capital for open source software adoption | 0.1509 |
| Discussion about the balance of human capital - taking Huawei technologies company as an example                                     | 0.1467 |
| Industry-specific human capital and wages: evidence from the business process outsourcing industry                                   | 0.1455 |
| A study on the new method to measure the increment of enterprise's human capital   | 0.1368 |
| Investing in core human capital and its effect on operating performance  | 0.1296 |
| The changing role of human resource management in an era of digital transformation   | 0.1287 |
| Editorial: human resource management in high-tech companies  | 0.1277 |

These results are strictly dependent on the topic words and the weights that we decided. The chosen words with the relative weights are:

- technology: 3
  - human: 3
  - intellectual: 1
  - employee: 1
  - information: 3
  - resource: 1
  - digital: 2
  - software: 3
  - high tech: 3
  - investment: 1
  - industry: 1
  - ict: 2
  - it: 2

For the first two publications we display the most frequent words in the abstract using the WordCloud(8) package.

Figure 3: WordClouds



## 9 PRISMA flow chart

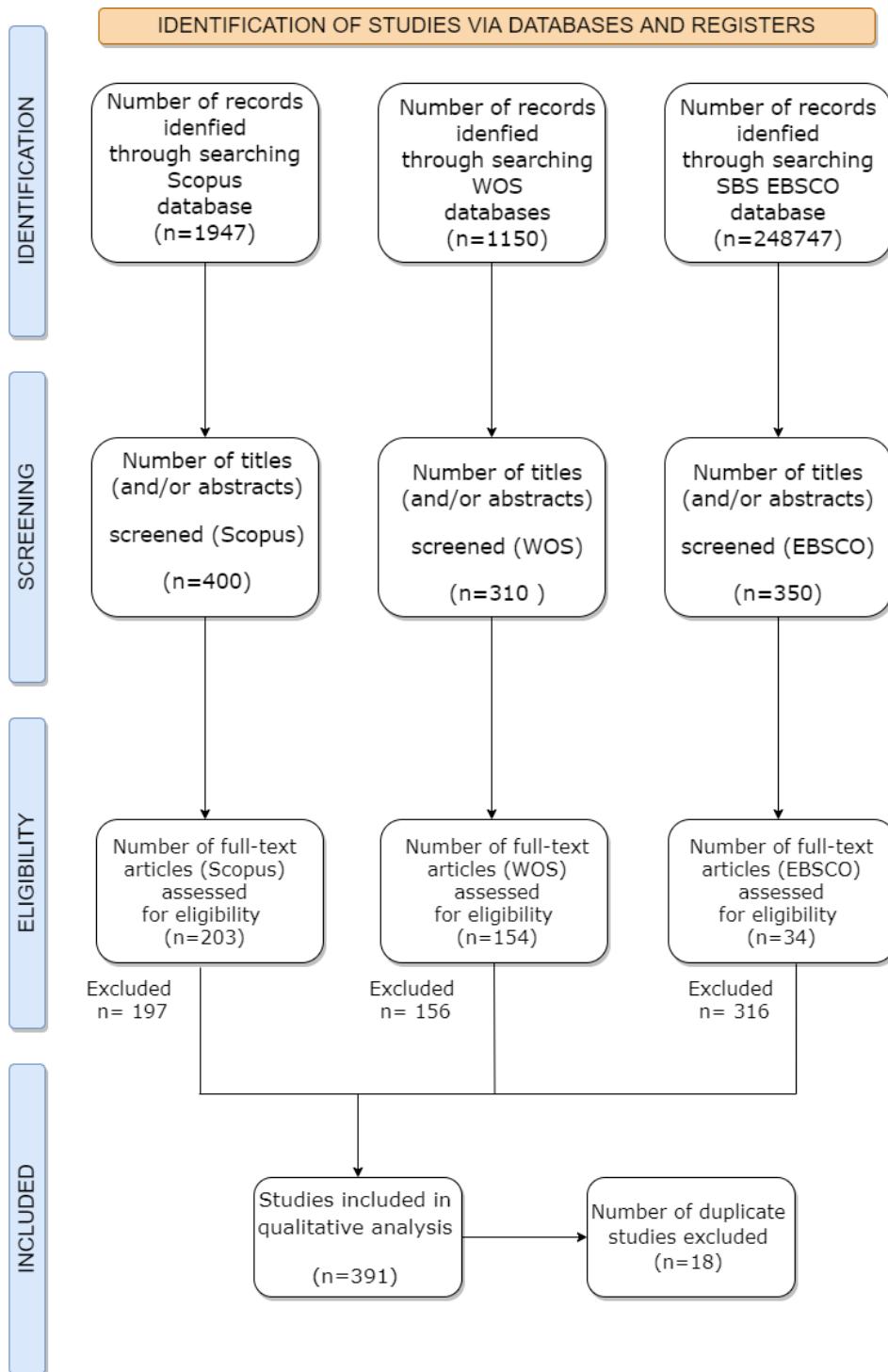


Figure 4: PRISMA flow chart

## 10 Conclusion

With this systematic review our group aimed to retrieve publications coherent with our research topic.

The first step has been to collect documents from a few important databases in the field of scientific publications, such as *Scopus*, *Web of Science* and *EBSCO*. This first retrieval was accomplished with queries formulated according to a specific schema (SPICE).

Afterwards, we have visualized with specific tools a few networks related to the single database results.

As a last step, we built a selection procedure based on the language processing on the abstract with the goal of retrieving the top-k documents.

All the steps in this systematic review are based on the Prisma Model described on the previous page.

To evaluate our work, it would be necessary a ground truth for the retrieved results.

However, we empirically observed that, among the most relevant publications retrieved, there are pertinent publications which can be considered to continue the analysis of the impact of human capital in the Hi-tech industry.

## References

- [1] D Tranfield, D Denyer and P Smart (2003) Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review, *British Journal of Management*, Vol. 14, 207–222;
- [2] How to do a systematic review, Alex Pollock and Eivind Berge, *International Journal of Stroke* 2018, Vol. 13(2) 138–156
- [3] Simone Di Leo, PRISMA Model In Action - Systematic review of university performance using intellectual capital, *Quantitative Models For Economic Analysis And Management Course Sapienza*, 15/04/2021
- [4] J. PT Higgins and S Green, (2008), *Cochrane Handbook for Systematic Reviews of Interventions; Systematic Reviews: Synthesis of Best Evidence for Clinical Decisions*
- [5] Manual for VOSviewer version 1.6.8, Nees Jan van Eck and Ludo Waltman, 27 April 2018
- [6] Pandas documentation: <https://pandas.pydata.org/pandas-docs/stable/>
- [7] NLTK documentation: <https://www.nltk.org/>
- [8] WordCloud for python documentation:[http://amueller.github.io/word\\_cloud/](http://amueller.github.io/word_cloud/)
- [9] code used for this report available at:  
<https://github.com/eugibaldo/Systematic-Review>