Findings Seminal Papers Using Data Mining Techniques

Alexander Báez Hernández¹, Debrayan Bravo Hidalgo²

¹Administrative Sciences Department, Universidad Central del Ecuador (UCE), Quito, Ecuador
²Department of Electrical Engineering and Electronics, Universidad San Francisco de Quito (USFQ), Quito, Ecuador
Email: albaez@uce.edu.ec, dbrayanbh@gmail.com

Abstract

The aim of this contribution is to show the detection of seminal papers using data mining techniques. To achieve the objective of this research, Rapidminer Studio software and its data mining tools are used, based on data created with information extracted from Google Scholar and Scopus, in three different areas of knowledge. In this process, other softwares such as Microsoft Excel and Publish or Perish are used. Comparing the results obtained for the searches in Knowledge Management, Entrepreneurship and Marketing, it was obtained that there is no marked similarity between the sets of articles that were obtained in Google Scholar and Scopus. The values for the Similarity Index remained below 0.52%, similar between Knowledge Management and Entrepreneurship but decreasing for Marketing. The detection of outliers using Data Mining techniques and in particular using Rapidminer, allowed to determine the seminals papers for the three search terms analyzed and allowed to characterize these in the space, in Google Scholar and Scopus. It was shown that the seminal articles can be different if Google Scholar or Scopus is used. The results suggest determining for other search terms whether the trend found is maintained or not.

Keywords
Seminals Papers, Knowledge Management, Entrepreneurship, Marketing, Outlier Detection, Data Mining, State of the Art

1. Introduction

Knowing the articles that laid the foundations of a specialty or a specific topic of research has been defined, for many years, as one of the essential objectives of a literature review (Hart, 1998). The literature review, necessary in any investigation, has been defined with fairness by (Webster & Watson, 2002) as the analysis
of the past as an essential preparation of the vision of the future that every good scientific article should contain. The realization of the so-called “state of the art” has become an essential step in the realization of research: Although it seems amazing the realization of these “states of art” was recognized almost from the very appearance of scientific journals, already many centuries (Sciences, 1823).

One of the main objectives in the realization of a state of the art is to identify those articles that have seated so much, the possible conceptual bases, as methodological of discipline, that is to say, those contributions that in fact “do not age” (Singer, 2009). It is usual, therefore, in the specialized literature, to find both: to determine those seminal articles published in a given journal (Parkinson et al., 2013), the role of one of these contributions in a particular discipline (Dolman, Miralles, & de Jeu, 2014), in a specific technique (Nash, Walker, Gidwani, & Ajuied, 2015; Nash, Walker, Lucas, & Ajuied, 2016) or the most important in a given branch of science (Riordon, Zubritsky, & Newman, 2000).

The importance of identifying the so-called seminal articles has been recognized as a de facto standard in the realization of a state of the art in the most dissimilar disciplines. To identify these articles of unquestionable significance in an investigation (Berkani, Hanifi, & Dahmani, 2020; Silva, Villa, & Cabrera, 2020), different alternatives have been proposed such as the use of collaborative models (Wang & Blei, 2011) and the use of personalized systems for the recommendation of the most relevant articles (Pera & Ng, 2011). Less studied has been the fact of how to identify these and their possible genealogy (Bae, Hwang, Kim, & Faloutsos, 2011, 2014). The fact is that the current researcher is faced with a quantity of information that does not do anything simply to find the most relevant jobs and this requires considerable time and effort (Alonso, Perez, & Hidalgo, 2016; Bravo Hidalgo & León González, 2018).

Within this problematic this contribution started from the investigative idea that the seminal articles are recognized as such, do not age, it is for two reasons:

1) They have been cited in a significant way, that is, they are recognized by the scientific community.

2) They remain valid for several years.

These two simple reasons should lead them to stand out as outliers in space:

\[ VY = f(C) \]

where \( VY \) is the Validity in Years of a given article, that is, the time elapsed from the publication of the article until the current date:

\( C \) is the number of appointments received during that period for the article in question.

Data mining offers different possibilities for data analysis (Berkhin, 2006) including different techniques (Bakar, Mohemad, Ahmad, & Deris, 2006; Buthong, Luangsodsai, & Sinapiromsaran, 2013) and algorithms for the detection of values atypical (Ramaswamy, Rastogi, & Shim, 2000). At the same time, different applications have been developed (Rangra & Bansal, 2014) that facilitate the use of data mining. Among these, the Rapidminer offers a whole set of possibilities for
the analysis of data (Amer & Goldstein, 2012; Jungermann, 2009) and in particular for the detection of outliers (Buthong et al., 2013). The outlier has long been defined (Barnett & Lewis, 1974) as an observation, or set of observations, that seems to be inconsistent with the data set under analysis.

This contribution was proposed from these considerations to determine if in the space \( VY = \mathcal{A}(C) \) could be distinguished the seminal articles as outliers using the possibility offered by the Rapidminer (https://rapidminer.com/) to classify them in said space. Another aspect that cannot be ignored is how the articles are determined and the number of citations received by each one. For this purpose, it was also proposed to explore in this research which was the coincidence in relation to the articles considered as seminal when using the Google Scholar (Martín-Martín, Orduna-Malea, Harzing, & Delgado López-Cózar, 2017) compared to another Database of wide recognition by the scientific world, such as Scopus (Burnham, 2006).

2. Material and Methods

To form the space \( VY = \mathcal{A}(C) \), we proceeded to search both Scopus and Google Scholar for the following terms in English, in the Title of the articles and for the period 1960-2019:

1) Knowledge management
2) Marketing
3) Entrepreneurship

The Publish or Perish (POP) tool (Harzing, 2007), which has been applied in different bibliometric studies (Harzing & Alakangas, 2016; Jacsó, 2009), was used to search Google Scholar.

For each of the search terms, the 990 most-cited articles were selected. These were exported to Excel according to the possibilities offered by both Scopus and POP. The Database is thus formed by the fields: Cites, Authors, Title, Year, and Validity that is calculated by subtracting the year of publication of the last year of rank for the search (2019).

In order to compare the similarity between the two sets of articles determined for each term, a Similarity Index (\( SL \)) was calculated from:

\[
SL = \frac{2C}{A_{Google\ Scholar} + B_{Scopus}}
\]

where:

\( SL \) is Similarity Index.

This \( SL \) reproduces the original idea of Sorensen, formulated many years ago to establish the similarity of groups of equal amplitude.

\( A_{Google\ Scholar} \) and \( B_{Scopus} \) are the number of Articles in each of the sets considered (990 for each).

\( C \) is the number of shared items of both sets. This number is easily calculated in Excel, if a formula is programmed that compares the coincidences for the two-column matrix (TitleGoogle Schlar, TitleScopus).

The detection of outliers was done using Rapidminer and the process scheme
that can be configured in this is shown in Figure 1.

The first Operator reads the file in Excel and processes the Cites and Validity fields, this was done for each search term and for each of the Bases used (Google Scholar and Scopus). The second identifies the Outliers in the data set. This allows you to specify both the number of neighbors ($k$), and the number of Outliers ($n$). To be able to compare the different search terms, these parameters were adjusted, after some preliminary tests, to the values:

$n = 10$
$k = 10$

The calculation of the distances between the values of $k$ was made using the Euclidean distances between these values. In practical terms, an attempt was made to answer the question: How to determine the 10 articles that can be considered seminal for each of the search terms analyzed?

3. Results

Table 1 below summarizes the Number of Cites, Years, Cites/Year and Cites/Paper as well as the Similarity Index for the three search terms considered. This information is of great importance for the purpose of characterizing the papers detected in the different databases used.

The Similarity Index remains similar between Knowledge Management and Entrepreneurship but decreases for Marketing.

![Figure 1](image.jpg) Process outlier detection in Rapidminer. The data used are those detected under the search criteria previously defined.

Table 1. Summary data. Google scholar matches with Scopus. Similarity index.

| Source         | Papers | Cites   | Years | Cites/Year | Cites/Paper | Coincidence | Title | SI |
|----------------|--------|---------|-------|------------|-------------|-------------|-------|----|
| Knowledge Management |        |         |       |            |             |             |       |    |
| Google Scholar  | 990    | 271,618 | 56    | 8770.90    | 274.36      | 509         | 0.51  |    |
| Scopus          | 990    | 107,574 | 35    | 3073.54    | 108.66      |             |       |    |
| Entrepreneurship|        |         |       |            |             |             |       |    |
| Google Scholar  | 990    | 494,796 | 56    | 8996.29    | 495.29      | 489         | 0.49  |    |
| Scopus          | 990    | 122,867 | 70    | 1755.24    | 124.11      |             |       |    |
| Marketing       |        |         |       |            |             |             |       |    |
| Google Scholar  | 990    | 1,002,743| 56  | 17,600.40  | 1005.23     | 265         | 0.27  |    |
| Scopus          | 990    | 174,304 | 52    | 3352.00    | 176.06      |             |       |    |
Analysis of Seed Articles

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable (Figure 2).

Table 2 presents the results for the SI for the case of articles determined as

![Figure 2](image_url)

**Figure 2.** Outliers in the space $VY_{Scopus} = f(C_{Scopus})$; knowledge management case.

Table 2. Seminals papers found: knowledge management, entrepreneurship and marketing.

| Knowledge Management          | Scopus                  | Google Scholar               |
|-------------------------------|-------------------------|------------------------------|
| Outlier | Cites | Author | Title | Years since publication | Reference | Outlier | Cites | Author | Title | Years since publication | Reference |
|-------|-------|--------|-------|-------------------------|-----------|-------|-------|--------|-------|-------------------------|-----------|
| 1     | true  | 4331.0 | Alavi M., Leidner D.E. | Knowledge management and knowledge management systems: Conceptual foundations and research issues | 18.0 | true | 10,629.0 | Alavi M., Leidner D.E. | Knowledge management and knowledge management systems: Conceptual foundations and research issues | 18.0 | (Alavi & Leidner, 2001) |
| 2     | true  | 1495.0 | Gold A.H., Malhotra A., Segars A.H. | Knowledge management: An organizational capabilities perspective | 18.0 | true | 4745.0 | TH Davenport, DW De Long... | Successful knowledge management projects | 21.0 | (Davenport, David, & Beers, 1998) |
Continued

3  true 1198.0
  Berkes F., Colding J., Folke C.
  Rediscovery of Traditional Ecological Knowledge as adaptive management
  Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review
  19.0  (Berkes, Colding, Folke, 2000)
  18.0  (Gold et al., 2001)
  23.0  (Sanchez & Mahoney, 1996)
  25.0  (Gomez-Perez, Fernandez-Lopez, & Corcho, 2006)

4  true 1153.0
  Tranfield D., Denyer D., Smart P.
  Modularity, flexibility, and knowledge management in product and organization design
  23.0  (Sanchez & Mahoney, 1996)
  19.0  (De Long & Fahey, 2000)
  19.0  (Shane & Venkataraman, 2000)
  19.0  (Shane & Venkataraman, 2000)

5  true 1118.0
  Sanchez R., Mahoney J.T.
  Knowledge management enablers, processes, and organizational performance: An integrative view and empirical examination
  A model of knowledge management and the N-form corporation
  16.0  (Lee & Choi, 2003)
  25.0  (Hedlund, 1994)
  25.0  (Hedlund, 1994)
  6.0  (Dalkir, 2013)

6  true 850.0
  De Long D.W., Fahey L.
  Diagnosing cultural barriers to knowledge management
  19.0  (De Long & Fahey, 2000)
  19.0  (De Long & Fahey, 2000)

7  true 835.0
  Lee H., Choi B.
  A general theory of entrepreneurship: The individual-opportunity nexus
  13.0  (Shane, 2003)
  4.0  (Kirzner, 2015)

8  true 833.0
  Hedlund G.
  Finance, entrepreneurship and growth
  26.0  (King & Levine, 1993)
  23.0  (Baumol, 1996)
| No. | Cites | Author(s) | Title | Years since publication | Reference |
|-----|-------|-----------|-------|--------------------------|-----------|
| 5   | 68.0  | Acs Z.J., Audretsch D.B., Lehmann E.E. | The knowledge spillover theory of entrepreneurship | 10.0 | (Acs, Braunerhjelm, Audretsch, & Carlsson, 2009) |
|     |       |           |       |                          | IA Timmons, S Spinelli, Y Tan | New venture creation: Entrepreneurship for the 21st century | 22.0 | (Timmons & Spinelli, 2004) |
|     | 39.0  | Zahra S.A. | Exploring the moderating effect of ownership and cohesion | 4.0 | (Zahra, 2012) |
| 6   | 37.0  | Fayolle A. | The dynamic of the entrepreneurial process | 9.0 | (Fayolle, 2007) |
|     |       |           |       |                          | SA Shane | A general theory of entrepreneurship: The individual-opportunity nexus | 13.0 | (Shane, 2003) |
| 7   | 37.0  | Baron R.A. | Some reciprocal benefits of closer links | 13.0 | (Baron, 2003) |
|     |       |           |       |                          | JG Covin, DP Slevin | A conceptual model of entrepreneurship as firm behavior | 25.0 | (Zahra, 1993) |
| 9   | 36.0  | Smallbone D., Welter F. | Entrepreneurship and institutional change in transition economies: The Commonwealth of Independent States, Central and Eastern Europe and China compared | 4.0 | (Smallbone & Welter, 2012) |
|     |       |           |       |                          | HH Stevenson, JC Jarillo | A paradigm of entrepreneurship: Entrepreneurial management | 9.0 | (Stevenson & Jarillo, 2007) |

### Marketing

| Outlier | Cites | Author(s) | Title | Years since publication | Reference |
|---------|-------|-----------|-------|--------------------------|-----------|
| 1       | 4403.0| Vargo S.L., Lusch R.F. | Evolving to a New Dominant Logic for Marketing | 12.0 | (Vargo & Lusch, 2004) |
|         |       |           |       |                          | P Kotler, D Gertner | Country as brand, product, and beyond: A place marketing and brand management perspective | 14.0 | (Kotler & Gertner, 2002) |
| 2       | 2280.0| Hoffman D.L., Novak T.P. | Marketing in hypermedia computer-mediated environments: Conceptual foundations | 20.0 | (Hoffman & Novak, 1996) |
|         |       |           |       |                          | P Kotler, G Armstrong | Principles of marketing | 6.0 | (Kotler & Armstrong, 2015) |
Outliers and that can be categorized as seminal using Google Scholar and Scopus and for the three search criteria used. The results obtained for the three search terms used are shown below in Table 2. In other words, this table identifies each of the detected documents as Outliers.

4. Conclusion

When comparing the results obtained for searches in Scopus and Google Scholar for Knowledge Management, Entrepreneurship and Marketing, it was obtained that there is no marked similarity between the sets of articles that were...
obtained in both cases. The values for the Similarity Index remained below 0.52%, similar between Knowledge Management and Entrepreneurship but decreasing for Marketing.

The detection of outliers using Data Mining techniques and in particular using Rapidminer, allowed to determine the seminal papers for the three search terms analyzed and allowed to characterize these in the space $VA = f(C)$ in Google Scholar and Scopus. It was shown that the seminal articles can be different if Google Scholar or Scopus is used. The results suggest determining for other search terms whether the trend found is maintained or not.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Acs, Z. J., Braunerhjelm, P., Audretsch, D. B., & Carlsson, B. (2009). The Knowledge Spillover Theory of Entrepreneurship. *Small Business Economics, 32*, 15-30. https://doi.org/10.1007/s11187-008-9157-3

Alavi, M., & Leidner, D. E. (2001). Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. *MIS Quarterly, 25*, 107-136. https://doi.org/10.2307/3250961

Alonso, J. A. G., Perez, Y. M., & Hidalgo, D. B. (2016). Empleo de indicadores bibliométricos para la realización de un estado del arte. Un enfoque práctico. *Revista Publicando, 3*, 81-97.

Amer, M., & Goldstein, M. (2012). Nearest-Neighbor and Clustering Based Anomaly Detection Algorithms for RapidMiner. In *the 3rd RapidMiner Community Meeting and Conference*.

Bae, D. H., Hwang, S. M., Kim, S. W., & Faloutsos, C. (2011). Constructing Seminal Paper Genealogy. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management* (pp. 2101-2104). https://doi.org/10.1145/2063576.2063900

Bae, D. H., Hwang, S. M., Kim, S. W., & Faloutsos, C. (2014). On Constructing Seminal Paper Genealogy. *IEEE Transactions on Cybernetics, 44*, 54-65. https://doi.org/10.1109/TCYB.2013.2246565

Bakar, Z. A., Mohemad, R., Ahmad, A., & Deris, M. M. (2006). A Comparative Study for Outlier Detection Techniques in Data Mining. In *IEEE Conference on Cybernetics and Intelligent Systems*. https://doi.org/10.1109/ICCIS.2006.252287

Barnett, V., & Lewis, T. (1974). *Outliers in Statistical Data*. Hoboken, NJ: Wiley.

Baron, R. A. (2003). Human Resource Management and Entrepreneurship: Some Reciprocal Benefits of Closer Links. *Human Resource Management Review, 13*, 253-256. https://doi.org/10.1016/S1053-4822(03)00016-0

Baumol, W. J. (1996). Entrepreneurship: Productive, Unproductive, and Destructive. *Journal of Business Venturing, 11*, 3-22. https://doi.org/10.1016/0883-9026(94)00014-X

Berkani, L., Hanifi, R., & Dahmani, H. (2020) Hybrid Recommendation of Articles in Scientific Social Networks Using Optimization and Multiview Clustering. In *3rd International Conference on Smart Applications and Data Analysis for Smart Cyber-Physical*
Systems (pp. 117-132, Vol. 1207). Berlin: Springer. https://doi.org/10.1007/978-3-030-45183-7_9

Berkes, F., Colding, J., & Folke, C. (2000). Rediscovery of Traditional Ecological Knowledge as Adaptive Management. Ecological Applications, 10, 1251-1262. https://doi.org/10.1890/1051-0761(2000)010[1251:ROTEKA]2.0.CO;2

Berkhin, P. (2006). A Survey of Clustering Data Mining Techniques Grouping Multidimensional Data (pp. 25-71). Berlin: Springer. https://doi.org/10.1007/3-540-28349-8_2

Berry, L. L. (1995). Relationship Marketing of Services—Growing Interest, Emerging Perspectives. Journal of the Academy of Marketing Science: Official Publication of the Academy of Marketing Science, 23, 236-245. https://doi.org/10.1177/009207039502300402

Bravo Hidalgo, D., & León González, J. L. (2018). Divulgación de la investigación científica en el Siglo XXI. Revista Universidad y Sociedad, 10, 88-97.

Burnham, J. F. (2006). Scopus Database: A Review. Biomedical Digital Libraries, 3, 1. https://doi.org/10.1186/1742-5581-3-1

Buthong, N., Luangsodsai, A., & Sinapiromsaran, K. (2013). Outlier Detection Score Based on Ordered Distance Difference. In Computer Science and Engineering Conference. https://doi.org/10.1109/ICSEC.2013.6694771

Churchill Jr., G. A. (1979). A Paradigm for Developing Better Measures of Marketing Constructs. Journal of Marketing Research, 16, 64-73. https://doi.org/10.1177/002224377901600110

Dalkir, K. (2013). Knowledge Management in Theory and Practice. Abingdon-on-Thames: Taylor and Francis. https://doi.org/10.4324/9780080547367

Davenport, T. H., David, W., & Beers, M. C. (1998). Successful Knowledge Management Projects. Sloan Management Review, 39, 43-58.

De Long, D. W., & Fahey, L. (2000). Diagnosing Cultural Barriers to Knowledge Management. Academy of Management Executive, 14, 113-127. https://doi.org/10.5465/ame.2000.3979820

Dolman, A. J., Miralles, D. G., & de Jeu, R. A. M. (2014). Fifty Years since Monteith’s 1965 Seminal Paper: The Emergence of Global Ecohydrology. Ecohydrology, 7, 897-902. https://doi.org/10.1002/eco.1505

Drucker, P. (2014). Innovation and Entrepreneurship. Abingdon-on-Thame: Routledge. https://doi.org/10.4324/9781315747453

Fayolle, A. (2007). Entrepreneurship and New Value Creation: The Dynamic of the Entrepreneurial Process. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511488689

Gold, A. H., Malhotra, A., & Segars, A. H. (2001). Knowledge Management: An Organizational Capabilities Perspective. Journal of Management Information Systems, 18, 185-214. https://doi.org/10.1080/07421222.2001.11045669

Gomez-Perez, A., Fernández-López, M., & Corcho, O. (2006). Ontological Engineering: with Examples from the Areas of Knowledge Management, e-Commerce and the Semantic Web. Berlin: Springer Science & Business Media.

Gronroos, C. (1984). A Service Quality Model and its Marketing Implications. European Journal of Marketing, 18, 36-44. https://doi.org/10.1108/EUM0000000004784

Hart, C. (1998). Doing a Literature Review: Releasing the Social Science Research Imagination. Thousand Oaks, CA: Sage Publications.

Harzing, A.-W. (2007). Publish or Perish.
Harzing, A.-W., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: A Longitudinal and Cross-Disciplinary Comparison. *Scientometrics, 106*, 787-804. [https://doi.org/10.1007/s11192-015-1798-9](https://doi.org/10.1007/s11192-015-1798-9)

Hedlund, G. (1994). A Model of Knowledge Management and the N-Form Corporation. *Strategic Management Journal, 15*, 73-90. [https://doi.org/10.1002/smj.4250151006](https://doi.org/10.1002/smj.4250151006)

Henseler, J., Ringle, C. M., & Sinkovics, R. R. (2009). The Use of Partial Least Squares Path Modeling in International Marketing. In *Advances in International Marketing* (pp. 277-319, Vol. 20). Bingley: Emerald Group Publishing Ltd. [https://doi.org/10.1108/S1474-7979(2009)0000020014](https://doi.org/10.1108/S1474-7979(2009)0000020014)

Hoffman, D. L., & Novak, T. P. (1996). Marketing in Hypermedia Computer-Mediated Environments: Conceptual Foundations. *Journal of Marketing, 60*, 50-68. [https://doi.org/10.2307/1251841](https://doi.org/10.2307/1251841)

Hunt, S. D., & Vitell, S. (1986). A General Theory of Marketing Ethics. *Journal of Macromarketing, 6*, 5-16. [https://doi.org/10.1177/027614678600600103](https://doi.org/10.1177/027614678600600103)

Jacsó, P. (2009). Calculating the h-Index and Other Bibliometric and Scientometric Indicators from Google Scholar with the Publish or Perish Software. *Online Information Review, 33*, 1189-1200. [https://doi.org/10.1108/14684520911011070](https://doi.org/10.1108/14684520911011070)

Jarvis, C. B., Mackenzie, S. B., Podsakoff, P. M., Giliatt, N., & Mee, J. F. (2003). A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research. *Journal of Consumer Research, 30*, 199-218. [https://doi.org/10.1086/376806](https://doi.org/10.1086/376806)

Jungermann, F. (2009). Information Extraction with Rapidminer. In *Proceedings of the GSCL Symposium: Sprachtechnologie und eHumanities*.

King, R. G., & Levine, R. (1993). Finance, Entrepreneurship and Growth. *Journal of Monetary Economics, 32*, 513-542. [https://doi.org/10.1016/0304-3932(93)90028-E](https://doi.org/10.1016/0304-3932(93)90028-E)

Kirzner, I. M. (2015). *Competition and Entrepreneurship*. Chicago, IL: University of Chicago Press.

Kotler, P., & Armstrong, G. (2015). *Principles of Marketing-Global Edition*. London: Pearson.

Kotler, P., & Gertner, D. (2002). Country as Brand, Product, and Beyond: A Place Marketing and Brand Management Perspective. *Journal of Brand Management, 9*, 249-261. [https://doi.org/10.1057/palgrave.bm.2540076](https://doi.org/10.1057/palgrave.bm.2540076)

Kotler, P., & Keller, K. (2000). *Marketing Administration*. Sao Paulo: Prentice Hall Publishers.

Kozinets, R. V. (2002). The Field behind the Screen: Using Netnography for Marketing Research in Online Communities. *Journal of Marketing Research, 39*, 61-72. [https://doi.org/10.1509/jmkr.39.1.61.18935](https://doi.org/10.1509/jmkr.39.1.61.18935)

Lee, H., & Choi, B. (2003). Knowledge Management Enablers, Processes, and Organizational Performance: An Integrative View and Empirical Examination. *Journal of Management Information Systems, 20*, 179-228. [https://doi.org/10.1080/07421222.2003.11045756](https://doi.org/10.1080/07421222.2003.11045756)

Malhotra, N. K. (2012). *Pesquisa de marketing: Uma orientação aplicada*. Porto Alegre: Bookman Editora.

Malhotra, N., Hall, J., Shaw, M., & Oppenheim, P. (2006). *Marketing Research: An Applied Orientation*. Melbourne: Pearson Education Australia.

Martin-Martín, A., Orduna-Malea, E., Harzing, A.-W., & Delgado López-Cózar, E. (2017). Can We Use Google Scholar to Identify Highly-Cited Documents? *Journal of Informetrics, 11*, 152-163. [https://doi.org/10.1016/j.joi.2016.11.008](https://doi.org/10.1016/j.joi.2016.11.008)
Miller, D. (1983). The Correlates of Entrepreneurship in Three Types of Firms. Management Science, 29, 770-791. https://doi.org/10.1287/mnsc.29.7.770

Morgan, R. M., & Hunt, S. D. (1994). The Commitment-Trust Theory of Relationship Marketing. Journal of Marketing, 58, 20-38. https://doi.org/10.1177/002224299405800302

Nash, W., Walker, R., Gidwani, S., & Ajuied, A. (2015). Seminal Papers in Hand and Wrist Surgery. Orthopaedics and Trauma, 29, 408-418. https://doi.org/10.1016/j.mporth.2015.08.002

Nash, W., Walker, R., Lucas, J., & Ajuied, A. (2016). Seminal Papers in Spinal Surgery. Orthopaedics and Trauma, 32, 263-274.

Palmatier, R. W., Dant, R. P., Grewal, D., & Evans, K. R. (2006). Factors Influencing the Effectiveness of Relationship Marketing: A Meta-Analysis. Journal of Marketing, 70, 136-153. https://doi.org/10.1509/jmkg.70.4.136

Percy, M. S., & Ng, Y.-K. (2011). A Personalized Recommendation System on Scholarly Publications. In Proceedings of the 20th ACM International Conference on Information and Knowledge Management. https://doi.org/10.1145/2063576.2063908

Ramaswamy, S., Rastogi, R., & Shim, K. (2000). Efficient Algorithms for Mining Outliers from Large Data Sets. SIGMOD Record (ACM Special Interest Group on Management of Data), 29, 427-438. https://doi.org/10.1145/342009.335437

Rangra, K., & Bansal, K. (2014). Comparative Study of Data Mining Tools. International Journal of Advanced Research in Computer Science and Software Engineering, 4, 2277.

Riordon, J., Zubritsky, E., & Newman, A. (2000). Top 10 Articles. Analytical Chemistry Looks at 10 Seminal Papers. Analytical Chemistry, 72, 324A-329A. https://doi.org/10.1021/ac002801q

Sanchez, R., & Mahoney, J. T. (1996). Modularity, Flexibility, and Knowledge Management in Product and Organization Design. Strategic Management Journal, 17, 63-76. https://doi.org/10.1002/smj.4250171107

Sciences, T. P. (1823). ART. I. A Comparative View of the State of Medical Science among the Ancients and Moderns, Its Revolutions in Different Periods of the World, and an Enumeration of Some of the Errors Which Check Its Progress. The Philadelphia Journal of the Medical and Physical Sciences, 7, 211-226.

Shane, S. (2003). A General Theory of Entrepreneurship: The Individual-Opportunity Nexus. Cheltenham: Edward Elgar Publishing. https://doi.org/10.4337/9781781007990

Shane, S., & Venkataraman, S. (2000). The Promise of Entrepreneurship as a Field of Research. Academy of Management Review, 25, 217-226. https://doi.org/10.5465/amr.2000.2791611

Silva, J., Villa, J. V., & Cabrera, D. (2020). An Intelligent Approach to Design and Development of Personalized Meta Search: Recommendation of Scientific Articles. In 16th International Conference on Distributed Computing and Artificial Intelligence, DCAI 2019 (pp. 99-106, Vol. 1003). Berlin: Springer Verlag. https://doi.org/10.1007/978-3-030-23887-2_12

Singer, W. (2009). Seminal Papers Don’t Age. Perception, 38, 799-802.

Smallbone, D., & Welte, F. (2012). Entrepreneurship and Institutional Change in Transition Economies: The Commonwealth of Independent States, Central and Eastern Eu-
rope and China Compared. *Entrepreneurship and Regional Development*, 24, 215-233. https://doi.org/10.1080/08985626.2012.670914

Stevenson, H. H., & Jarillo, J. C. (2007). A Paradigm of Entrepreneurship: Entrepreneurial Management. In *Entrepreneurship: Concepts, Theory and Perspective* (pp. 155-170). Berlin: Springer. https://doi.org/10.1007/3-540-48543-0_7

Tax, S. S., Brown, S. W., & Chandrashekaran, M. (1998). Customer Evaluations of Service Complaint Experiences: Implications for Relationship Marketing. *Journal of Marketing*, 62, 60-76. https://doi.org/10.1177/002224299806200205

Timmons, J. A., & Spinelli, S. (2004). *New Venture Creation: Entrepreneurship for the 21st Century* (Vol. 6). New York: McGraw-Hill.

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review. *British Journal of Management*, 14, 207-222. https://doi.org/10.1111/1467-8551.00375

Vargo, S. L., & Lusch, R. F. (2004). Evolving to a New Dominant Logic for Marketing. *Journal of Marketing*, 68, 1-17. https://doi.org/10.1509/jmkt.68.1.1.24036

Wang, C., & Blei, D. M. (2011). Collaborative Topic Modeling for Recommending Scientific Articles. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. https://doi.org/10.1145/2020408.2020480

Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, 26, 13-23.

Wilson, A., Zeithaml, V. A., Bitner, M. J., & Gremler, D. D. (2012). *Services Marketing: Integrating Customer Focus across the Firm*. New York: McGraw Hill.

Zahra, S. A. (1993). A Conceptual Model of Entrepreneurship as Firm Behavior: A Critique and Extension. *Entrepreneurship Theory and Practice*, 17, 5-21. https://doi.org/10.1177/104225879301700401

Zahra, S. A. (2012). Organizational Learning and Entrepreneurship in Family Firms: Exploring the Moderating Effect of Ownership and Cohesion. *Small Business Economics*, 38, 51-65. https://doi.org/10.1007/s11187-010-9266-7