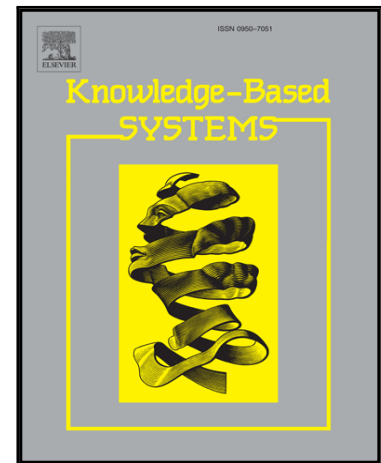


## Accepted Manuscript

Learning the evolution of disciplines from scientific literature: A functional clustering approach to normalized keyword count trajectories

Matilde Trevisani, Arjuna Tuzzi

PII: S0950-7051(18)30051-0  
DOI: [10.1016/j.knosys.2018.01.035](https://doi.org/10.1016/j.knosys.2018.01.035)  
Reference: KNOSYS 4208



To appear in: *Knowledge-Based Systems*

Received date: 14 February 2017  
Revised date: 28 January 2018  
Accepted date: 31 January 2018

Please cite this article as: Matilde Trevisani, Arjuna Tuzzi, Learning the evolution of disciplines from scientific literature: A functional clustering approach to normalized keyword count trajectories, *Knowledge-Based Systems* (2018), doi: [10.1016/j.knosys.2018.01.035](https://doi.org/10.1016/j.knosys.2018.01.035)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# Learning the evolution of disciplines from scientific literature: A functional clustering approach to normalized keyword count trajectories

Matilde Trevisani<sup>a,\*</sup>, Arjuna Tuzzi<sup>b</sup>

<sup>a</sup>*Department of Economics, Business, Mathematics and Statistics (DEAMS), University of Trieste, Via Tigor 22, 34124 Trieste (Italy)*

<sup>b</sup>*Department of Philosophy, Sociology, Education and Applied Psychology (FISPPA), University of Padova, Via M. Cesarotti 10/12, 35123 Padova (Italy)*

---

## Abstract

The growing availability of large diachronic corpora of scientific literature offers the opportunity of reading the temporal evolution of concepts, methods and applications, i.e., the history of disciplines involved in the strand under investigation. After a retrieval process of the most relevant keywords, bag-of-words approaches produce words  $\times$  time-points contingency tables, i.e. the frequencies of each word in the set of texts grouped by time-points. Through the analysis of word counts over the observed period of time, main purpose of the study is, after reconstructing the "life-cycle" of words, clustering words that have similar life-cycles and, thus, detecting prototypical or exemplary temporal patterns. Unveiling such relevant and (through expert opinion) meaningful inner dynamics enables us to trace a historical narrative of the discipline of interest. However, different history readings are possible depending on the type of data normalization, which is needed to account for the fluctuating size of texts across time and the general problems of data sparsity and strong asymmetry. This study proposes a methodology consisting of (1) a stepwise information retrieval procedure for keywords' selection and (2) a functional clustering two-stage approach for statistical learning. Moreover, a sample of possible normalizations of word

---

\*Corresponding author

*Email addresses:* [matilde.trevisani@deams.units.it](mailto:matilde.trevisani@deams.units.it) (Matilde Trevisani),  
[arjuna.tuzzi@unipd.it](mailto:arjuna.tuzzi@unipd.it) (Arjuna Tuzzi)

Download English Version:

<https://daneshyari.com/en/article/6861601>

Download Persian Version:

<https://daneshyari.com/article/6861601>

[Daneshyari.com](https://daneshyari.com)