



# Open Research Online

---

The Open University's repository of research publications and other research outputs

## How do some concepts vanish over time?

### Conference or Workshop Item

How to cite:

Pavel, Gabriela (2009). How do some concepts vanish over time? In: The 18th Annual Belgian-Dutch Conference on Machine Learning (Benelearn 09), 18-19 May 2009, Tilburg, The Netherlands.

For guidance on citations see [FAQs](#).

© 2009 Tilbury centre for Creative Computing; The Author

Version: Version of Record

Link(s) to article on publisher's website:

[http://benelearn09.uvt.nl/Proceedings\\_Benelearn\\_09.pdf#page=97](http://benelearn09.uvt.nl/Proceedings_Benelearn_09.pdf#page=97)

---

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

---

[oro.open.ac.uk](http://oro.open.ac.uk)

# How do some concepts vanish over time?

**Gabriela Pavel**

Knowledge Media Institute, The Open University

Milton Keynes, UK

g.pavel@open.ac.uk

## Abstract

This paper presents the current stage of my PhD research focused on the use of machine learning in supporting the human learn from examples. I present here an approach to answer how some concepts change their contexts in time, using two techniques suitable for indexing and data mining: latent semantic indexing (LSI) and the APRIORI algorithm.

## 1 Credits

This work is sustained by the educational project, SILVER, whose purpose is to use images and build technologies that will help pupils learn from examples. The project provides access to a large annotated collection of images, the Bridgeman Art Library of London<sup>1</sup>, from which can be selected the words representing the concepts of interest in the human learning process.

## 2 Introduction

In this paper I present how data mining techniques can be applied to help teachers select examples relevant for a given topic. The human learning process considered here contains three steps. 1) The teacher defines a topic and selects a target concept (concept to be learned during the lesson). For instance, in teaching about *citizenship* topic, teachers might need to underline the concept of *women's public/private roles in the society*. These roles represent changing concepts, i.e. they have different roles associated to them, roles like *governess* and *mother*. 2) Then the teacher selects the concepts that reflect the target concept (a descriptive vocabulary is introduced). I am interested to use machine learning to suggest similarities among concepts, because the learning curricula for the key-stage 3 (pupils of 11-14 years old) mentions two important learning steps: a) the learning of a concept in relation with others (e.g. learn about the public/private role considering particular roles) and b) the understanding of a concept in different periods of time. The focus is on the use of a set of words

provided by selected teachers within the SILVER project. Words illustrate different women roles. 3) The teacher asks the pupils to group concepts according to their similarities. Students could receive suggested combinations of roles and be asked to assign a meaning to the combinations.

Given the problem, I have to answer two questions: 1) How can machine learning techniques be used to help the teacher in the selection of examples? 2) What roles associations to suggest to the student? I think that data mining is suitable to do this for two reasons: first of all it selects similarities between concepts representing roles; second, I observed that it reflects the contexts changes (conceptual changes over time).

In order to start answering the above questions, I looked for corpora that reflect semantic relationships between the concepts of interest. For this purpose I analyzed two collections of text documents, using two data mining methods: latent semantic indexing (LSI) - to consider similarities between concepts (reflected in the cosine values) - and the APRIORI algorithm - to detect the contexts for these concepts. I describe shortly the resources used, then the methods used, and finally some conclusions and future plans.

## 3 Case studies

I built a case study with the objective to investigate the changes of women roles, wrt. other roles, between the Victorian period and present. I used two dataset with texts extracted from two collections of online-accessible documents. They are: 1) a collection of documents from the Victorian Women Writers Project (contains transcriptions of works of British women writers from the 19<sup>th</sup> century, where roles begin to change; 2) a selection of articles from the Guardian, browsing for texts that contain the concepts of interest. Databases are described in table 1.

Collection	Size	Period
VWWP	184 files	19 <sup>th</sup> century
Guardian	312 files	21 <sup>th</sup> century

Table 1. Case studies (collections used)

<sup>1</sup> <http://www.bridgemaneducation.com>

Both datasets were selected so that the texts describe 15 roles women played in the Victorian period. By thus, the considered roles don't include modern roles, such as *pilot*.

With the built resources I made two experiments in order to understand how to associate roles. For the experiments I used existing implementations of LSI and the Apriori algorithm.

### 3.1 Using LSI to differentiate among terms

Latent semantic indexing method (LSI) determines the concepts statistically using the singular value decomposition (SVD) method for decomposing words across documents (Berry et al., 1995). I used the cosine values to compute the conceptual indexes for the two datasets. I used the R implementation of the LSI (Wild, 2009).

Out of the 15 concepts (roles), I created 105 pairs of concepts to characterize inter-concepts relationships. Relationships were measured as cosine similarities values of the pairs wrt. the corpora of the two periods. Results are shown in fig.1. Integers on X axis are associated with pairs, while Y axis represents cosine values for each pair.

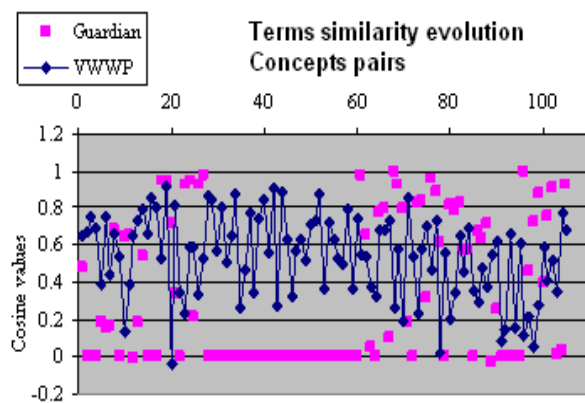


Figure 1: Terms similarities for the case studies

I observed that that some roles associations seem to vanish in time (disappear or loose importance in documents from a different time period, such as combinations of *nursemaid*).

According to the cosine values in both periods, there are four possible situations: some roles keep a significant low value, or on the contrary a non-interesting high value, or they modify their values from one source to the other (see table 3).

Roles pairs	S1	S2	Change
( <i>doctor, officer</i> )	0.37	0.92	LH
( <i>nursemaid, scholar</i> )	0.73	0	HL
( <i>mayor, officer</i> )	0.75	0.76	HH
( <i>officer, scholar</i> )	0.36	0.46	HL

Table 3: Example of roles changes

From pairs changes I can deduce the changes of concepts. These examples of changes will be brought to the students and they will be asked to reason about these temporal conceptual changes.

### 3.2. Building contexts using APRIORI

I created association rules to build contexts of the changing concepts (a context is a rule for co-occurrences of terms) using the APRIORI algorithm (Boden, 2001). Item-sets considered here contain codifications of text documents; a document is codified by the indexes of concepts that occur in it. The set of rules being too large, I considered the minimization of rules which was not very interesting for this case. I will have to look for a strategy to compare the rules extracted from the two sources in order to understand the temporal changes.

## 4 Future work

I will have to introduce the extracted information in the learning scenario in two ways: 1) associating images to distinct periods, according to their current annotations, 2) suggesting annotations to images as associations of roles according to the period of the images. I will also have to evaluate the learning process by means of questionnaires. I hope that my approach will support pupils learn concepts from images.

### Acknowledgments

Thank you to my supervisors dr. Zdenek Zdrahal and dr. Paul Mulholland for their professional advice and support.

### References

Berry, M. W. et al. 1995. *Using linear algebra for intelligent information retrieval*, SIAM Review, vol. 37, pp. 573-595

Boden, F. 2001. *A fast APRIORI implementation*, IEEE ICDM Workshop on Frequent Itemset Mining Implementations (FIMI'03), USA, 2003.

Wild, F. 2009. *LSI package for R*, <http://cran.at-r-project.org/web/packages/LSI/index.html>

*The Guardian* (consulted in 2009), <http://browse.guardian.co.uk>

*Victorian Women Writers Project* (consulted in 2009), <http://www.indiana.edu/~letrs/vwwp/>