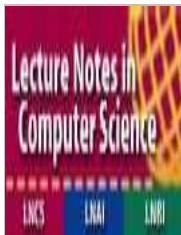


> [Home](#) / [Publication](#) / [Volume](#) /

## Chapter



### Lecture Notes in Computer Science

Publisher: Springer Berlin / Heidelberg

ISSN: 0302-9743

Subject: [Computer Science](#)

**Volume 3878 / 2006**

Title: Computational Linguistics and Intelligent Text Processing: 7th International Conference, CICLing 2006, Mexico City, Mexico, February 19-25, 2006. Proceedings

Editors: Alexander Gelbukh

ISBN: 3-540-32205-1

DOI: 10.1007/11671299

**Chapter:** pp. 536 - 546

DOI: 10.1007/11671299\_55

[Previous chapter](#)

[Next chapter](#)

**Export Citation:** [RIS](#) | [Text](#)

**Linking Options**

**Send this article to**

Go

**Quick Search**

Search within this publication...

For:

- Search Title/Abstract Only
- Search Author
- Search Fulltext
- Search DOI

Search

**Full Text Secured**

The full text of this article is secured to subscribers. To gain access, you may:

- Add this item to your shopping cart for purchase later.

- Purchase this item now.

- Log in to verify access.

### Text Classification

## Clustering Abstracts of Scientific Texts Using the Transition Point Technique

David Pinto<sup>1, 2</sup> , Héctor Jiménez-Salazar<sup>1</sup>  and Paolo Rosso<sup>2</sup> 

(1) Faculty of Computer Science, BUAP, Puebla 72570, Ciudad Universitaria, Mexico

(2) Department of Information Systems and Computation, UPV, Valencia 46022, Camino de Vera s/n, Spain

### Abstract

Free access to scientific papers in major digital libraries and other web repositories is limited to only their abstracts. Current keyword-based techniques fail on narrow domain-oriented libraries, e.g., those containing only documents on high energy physics like those of the *hep-ex* collection of CERN. We propose a simple procedure to cluster abstracts which consists in applying the transition point technique during the term selection process. This technique uses the mid-frequency terms to index the documents due to the fact that they have a high semantic content. In the experiments we have carried out, the transition point approach has been compared with well known unsupervised term selection techniques. Transition point technique shown that it is possible to obtain a better performance than traditional methods. Moreover, we propose an approach to analyse the stability of transition point term selection method.

This work was partially supported by BUAP-VIEP 3/G/ING/05, R2D2 (CICYT TIC2003-07158-C04-03), ICT EU-India (ALA/95/23/2003/077-054), and Generalitat Valenciana Grant (CTESIN/2005/012).



**David Pinto**

Email: [dpinto@dsic.upv.es](mailto:dpinto@dsic.upv.es)



**Héctor Jiménez-Salazar**

Email: [hgimenezs@gmail.com](mailto:hgimenezs@gmail.com)



**Paolo Rosso**

Email: [proso@dsic.upv.es](mailto:proso@dsic.upv.es)

*The references of this article are secured to subscribers.*

[Frequently asked questions](#) | [General information on journals and books](#)

---

© Springer. Part of [Springer Science+Business Media](#) | [Privacy](#), [Disclaimer](#), [Terms and Conditions](#), © [Copyright Information](#)

Remote Address: 148.228.22.73 • Server: MPWEB19

HTTP User Agent: Mozilla/5.0 (X11; U; Linux i686; en-US; rv:1.7.12) Gecko/20050920 Firefox/1.0.7 SUSE/1.0.7-0.1