

## List of References

1. Bagdanov A. (2004), *Style Characterization of Machine Printed Texts*, PhD dissertation submitted for the degree of Doctor of Philosophy, University of Amsterdam.
2. Bagdanov A. and Worring M. (2001a), Content-free Document Genre Classification using First Order Random Graphs, *Proceedings of the 7th Annual Conference of the Advanced School for Computing and Imaging*.
3. Bagdanov A. and Worring M. (2001a), Fine-Grained Document Genre Classification Using First Order Random Graphs, *Proceedings of the Sixth International Conference on Document Analysis and Recognition*.
4. Beaudouin V., Fleury S., Habert B., Illouz G., Licoppe C., Pasquier M. (2001a), TyPWeb: décrire la Toile pour mieux comprendre les parcours, *Colloque International sur les Usages et les Services des Télécommunications, e-Usages*.
5. Beaudouin V., Fleury S., Habert B., Illouz G., Licoppe C., Pasquier M. (2001b), Traits textuels, structurels et présentationnels pour typer les sites web personnels et marchands, available online.
6. Biber D. (1988), *Variations across speech and writing*, Cambridge University Press.
7. Biber D. (1989), A typology of English texts, *Linguistics*, Vol. 27, 3-43.
8. Biber D. (1993), The Multi-Dimensional Approach to Linguistic Analyses of Genre Variation: An Overview of Methodology and Findings, *Computers and the Humanities*, Vol. 26, 331-345.
9. Biber D. (1995), *Dimensions of register variation*, Cambridge University Press, Cambridge.
10. Biber D. (2004a), Conversation text types: A multi-dimensional analysis, *7es Journées internationales d'Analyse statistique des Données Textuelles*.
11. Biber D. (2004b), Towards a typology of web registers: A multi-dimensional analysis. Invited lecture, Conference on Corpus Linguistics: Perspectives for the future. University of Heidelberg, Germany, October 2004.
12. Biber D., Davies M., Jones J. and Tracy N. (2005), Register variation in Spanish: A multi-dimensional comparison with English, Somali, and Korean, *AAACL-6/ICAME-26*.
13. Boese E. (2005a), *Stereotyping the Web: Genre Classification of Web Documents*, M.S. Thesis, Computer Science Department, Colorado State University, Fort Collins, CO.
14. Boese E. (2005b), Genre Classification of Web Documents. Poster at *AAAI*.
15. Boese E. and Howe A. (2005), Effects of Web Document Evolution on Genre Classification, *CIKM 2005*.

**Marina Santini**

*From Biberian Text Types to Genres of Web Pages:  
An Overview of Studies on Automatic Genre Identification*

Toulouse, 5 et 6 octobre 2006

<http://w3.univ-tlse2.fr/erss/textes/seminaires/sc2006/sc2006.html>

16. Crowston K. and Kwasnik B. (2004), A Framework for Creating a Facetted Classification for Genres: Addressing Issues of Multidimensionality, *Proceedings of the 37th Hawaii International Conference on System Science*.
17. Crowston K. and Williams M. (1997), Reproduced and Emergent Genres of Communication on the World-Wide Web, *Proceedings of the 30th Hawaii International Conference on System Sciences*.
18. Dewdney N., Vaness-Dikema C. and Macmillan R. (2001), The form is the Substance: Classification of Genres in Text, *Proceeding of ACL*.
19. Dillon A. and Gushrowski B. (2000), Genres and the Web: is the personal home page the first uniquely digital genre?, *Journal of the American Society for Information Science*, Vol. 51(2).
20. Dillon A., Vaughan M. (1997), It's the journey and the destination: Shape and the emergent property of genre in evaluating digital documents, *New Review of Multimedia and Hypermedia*, Vol. 3, 91-106.
21. Dimitrova M., Finn A., Kushmerick N. and Smyth B. (2002), Web Genre Visualization, submitted to the *Conference on Human Factors in Computing Systems*.
22. Dimitrova M., Kushmerick N., Radeva P. and Villanueva J.(2003), User Assessment Of A Visual Web Genre Classifier, available online.
23. Finn A. and Kushmerick N. (2003), Learning to classify documents according to genre, *Workshop on Computational Approaches to Text Style and Synthesis*.
24. Finn A. and Kushmerick N. (2006). Learning to classify documents according to genre, *Journal of the American Society for Information Science and Technology*, Special Issue on Computational Analysis of Style, volume 7, number 5, march 2006.
25. Finn A., Kushmerick N. and Smyth B. (2002), Genre classification and domain transfer for information filtering, *Proceedings of European Colloquium for Information Retrieval Research*.
26. Finn A., Kushmerick N., and Smyth B. (2001), Fact or fiction: Content classification for digital libraries, *Joint DELOS-NSF Workshop on Personalisation and Recommender Systems in Digital Libraries*.
27. Folch H., Heiden S., Haber B., Fleury S., Illouz G., Lafon P., Nioche J. and Prévost S. (2000), TyPText: Inductive typological text classification by multivariate statistical analysis for NLP systems tuning/evaluation, *Second International Conference on Language Resources and Evaluation*.
28. Habert B., Illouz G., Fleury S., Folch H., Heiden S., Prévost S. (2000), Profilage de textes: cadre de travail et expérience", *5es Journées Internationales d'Analyse Statistique des Données Textuelles*.
29. Illouz G., Habert B. (2002), TyPWeb: Typologie et Profilage de sites Web, available online.

**Marina Santini**

*From Biberian Text Types to Genres of Web Pages:  
An Overview of Studies on Automatic Genre Identification*

Toulouse, 5 et 6 octobre 2006

<http://w3.univ-tlse2.fr/erss/textes/seminaires/sc2006/sc2006.html>

30. Illouz G., Habert B., Fleury, S., Folch H., Heiden S. and Lafon S. (1999), *Maîtriser les déluges de données hétérogènes, Atelier Thématique TALN 1999.*
31. Illouz G., Habert B., Folch H., Heiden S., Fleury Serge, Lafon S. and Prévost S. (2000), *TyPText: Generic features for Text Profiler, Content-Based Multimedia Information Access.*
32. Johannesson E. and Wallström C. (1999), *Automatic Analysis and Visualization of Stylistic Genres, 22nd IRIS Conference (Information Systems Research Seminar In Scandinavia).*
33. Karlgren J. (2000), *Stylistic Experiments for Information Retrieval*, Thesis submitted for the degree of Doctor of Philosophy, Stockholm University, Sweden.
34. Karlgren J. and Cutting D. (1994), *Recognizing Text Genre with Simple Metrics Using Discriminant Analysis, Proceedings of COLING 94.*
35. Karlgren J., Bretan I., Dewe J., Hallberg A. and Wolkert N. (1998a), *Iterative Information Retrieval Using Fast Clustering and Usage-Specific Genres, EIGHTH DELOS WORKSHOP, User Interface in Digital Libraries, Stockholm, Sweden, 21-23 October 1998.*
36. Kennedy A. and Shepherd M. (2005), *Automatic Identification of Home Pages on the Web, Proceedings of the 38th Annual Hawaii International Conference on System Sciences.*
37. Kessler B., Numberg G. and Shütze H. (1997), *Automatic Detection of Text Genre, Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics.*
38. Klavans J. and Kan M. (1998), *Role of Verbs in Document Analysis, Proceeding of COLING-ACL 1998.*
39. Lee D. (1999), *Modelling variation in spoken and written language: The multi-dimensional approach revisited*, Thesis submitted for the degree of Doctor of Philosophy, Lancaster University, UK. It is scheduled to be published by Routledge by the end of 2005/early 2006 (monograph series in corpus-based linguistics).
40. Lee D. (2001), *Genres, Registers, Text types, Domains, and Styles: Clarifying the concepts and navigating a path through the BNC Jungle, Language Learning and Technology, Vol. 5, Num. 3, pp. 37-72.*
41. Lee Y. and Myaeng S. (2002), *Text Genre Classification with Genre-Revealing and Subject-Revealing Features, Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval.*
42. Lee Y. and Myaeng S. (2004), *Automatic Identification of Text Genres and Their Roles in Subject-Based Categorization, Proceedings of the 37th Hawaii International Conference on System Science.*

**Marina Santini**

*From Biberian Text Types to Genres of Web Pages:  
An Overview of Studies on Automatic Genre Identification*  
Toulouse, 5 et 6 octobre 2006

<http://w3.univ-tlse2.fr/erss/textes/seminaires/sc2006/sc2006.html>

43. Lim C., Lee K. and Kim G.(2005), Automatic Genre Detection of Web Documents, Su K., Tsujii J., Lee J., Kwong O. Y. (eds.) *Natural Language Processing – IJCNLP 2004*, Springer, Berlin Heidelberg.
44. Mehler A. and Gleim R. (2006). The Net for the Graphs: Towards Webgenre Representation for Corpus. in Baroni M. and Bernardini S. (eds), *WaCky! Working Papers on the Web as Corpus*. pp. 191-334. GEDIT, Bologna (Italy).
45. Meyer zu Eissen S. and Stein B. (2004), Genre Classification of Web Pages: User Study and Feasibility Analysis, in Biundo S., Fruhwirth T. and Palm G. (eds.), *KI 2004: Advances in Artificial Intelligence*, Springer, Berlin Hedelberg New York, p. 256-269.
46. Nakamura, J. (1993), Statistical Methods and Large Corpora - A New Tool for Describing Text Type, in Baker M., Francis G., Tognini-Bonelli E. (eds.), *Text and Technology*, J. Benjamins Publishing Company, Philadelphia - Amsterdam, pp. 291-312.
47. Pery-Woodley M.-P and Rebeyrolle J., Domain and Genre in sublanguage text: definitional microtexts in three corpora, *First International Conference On Language Resources And Evaluation*.
48. Poudat C. and Cleuziou G. (2003), Genre and Domain Processing in and Information Retrieval Perspective, Cueva Lovelle J. M. et al. (eds.) *ICWE 2003*, Springer-Verlag, Berlin Heidelberg, 399-402.
49. Rauber A. and Müller-Kögler A. (2001), Integrating Automatic Genre Analysis into Digital Libraries, *ACM/IEEE joint Conference on Digital Libraries 2001*, Roanoke, USA.
50. Rehm G. (2002), Towards Automatic Web Genre Identification. A corpus-based approach in the Domain of Academia by Example of the Academic's Personal Homepage, *Proceedings of the 35th Hawaii International Conference on System Sciences*.
51. Rehm G. (2005), Language-Independent Text Parsing of Arbitrary HTML-Documents. Towards a Foundation for Web Genre Identification, *LDV Forum 2005*, Vol. 20(2), 53-74.
52. Rehm G (2006), Hypertext Types and Markup Languages, Metzling,D. and Witt A. (eds.) *Linguistic Modelling of Information and Markup Languages*, Springer, 2006.
53. Rosso M. (2005), *Using Genre to Improve Web Search*, PhD dissertation submitted for the degree of Doctor of Philosophy, University of North Carolina, Chapel Hill, USA.
54. Roussinov D., Crowston K., Nilan M., Kwasnik B., Cai J. and Liu X. (2001), Genre Based Navigation on the Web, *Proceedings of the 34th Hawaii International Conference on System Sciences*.
55. Santini M. (Forthcoming), *Automatic Identification of Genre of Web Pages*. PhD Thesis, University of Brighton, UK.
56. Shepherd M. and Watters C. (2004), Identifying Web Genre: Hitting a Moving Target, *Proceedings of the WWW2004 Conference. Workshop on Measuring Web Search Effectiveness: The User Perspective*.

**Marina Santini**

*From Biberian Text Types to Genres of Web Pages:  
An Overview of Studies on Automatic Genre Identification*

Toulouse, 5 et 6 octobre 2006

<<http://w3.univ-tlse2.fr/erss/textes/seminaires/sc2006/sc2006.html>>

57. Shepherd M. and Watters C. (1998), The Evolution of Cybergenre, *Proceedings of the 31st Hawaii International Conference on System Sciences*.
58. Shepherd M., Watters C. and Kennedy A. (2004), Cybergenre: Automatic Identification of Home Pages on the Web, *Journal of Web Engineering*, Vol. 3(3-4), 236-251.
59. Sigley R. (1997), Text Categories and Where You can Stick Them: A Crude Formality Index, *International Journal of Corpus Linguistics*, Vol. 2 No. 2, pp. 199-237.
60. Stamatatos E., Fakotakis N. and Kokkinakis G. (2000), Text Genre Detection Using Common Word Frequencies, *Proceeding of the COLING 2000*.
61. Stamatatos E., Fakotakis N. and Kokkinakis G. (2001), Automatic Text Categorization in Terms of Genre and Author, *Computational Linguistics*, Vol. 26(4), 471-495.
62. Takahashi K. (1997), A Study of Text Typology: Multi-feature and multi-dimensional analyses, *UCREL Technical Paper*, University of Lancaster.
63. Toms E. and Campbell D. (1999a). Genre as interface metaphor: Exploiting form and function in digital environments, *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences*.
64. Toms E. and Campbell D. (1999b). Utilizing information 'shape' as an interface metaphor based on genre, *Proceedings of the 27th Annual Conference of the Canadian Association for Information Science*.
65. Watters C. and Shepherd M. (1997), The Role of Genre in the Evolution of Interfaces for the Internet, *11th Annual Canadian Internet Conference*, NET 97.
66. Wikberg K. (1993), Verbs as indicators of text type and/or style: Some observations on the LOB corpus, in Souter C., Atwell E. (eds.), *Corpus-based Computational Linguistics*, Rodopi, Amsterdam–Atlanta, pp. 127-145.
67. Yin L. and Power R. (2006), Adapting Naive Bayes Classifier to Rank Procedural Texts, *EACL 2006 Student Session*.