

## Christoph Bussler

---

**From:** dbworld-bounces@cs.wisc.edu on behalf of Marenglen Biba [marenglenbiba@unyt.edu.al]  
**Sent:** Friday, June 18, 2010 6:41 AM  
**To:** dbworld@cs.wisc.edu  
**Subject:** [Dbworld] Call for Chapters

CALL FOR CHAPTERS

Chapter Proposals Submission Deadline: 20/07/2010 Full Chapters Due: 30/10/2010

LEARNING STRUCTURE AND SCHEMAS FROM DOCUMENTS

A book edited by:

Dr Marenglen Biba, University of New York Tirana, Albania Dr Fatos Xhafa, University of London (Birkbeck), United Kingdom To be published in the "Studies in Computational Intelligence" book series, Springer (2011) <http://www.marenglenbiba.net/cfc/>

Introduction

The rapidly growing volume of available digital documents of various formats and the possibility to access these through internet-based technologies, have led to the necessity to develop solid methods to properly organize and structure documents in large digital libraries and repositories. Specifically, since the extremely large volumes make it impossible to manually organize such documents and since most of the documents exist in an unstructured form and do not follow any schemas, most of the efforts in this direction are dedicated to automatically infer structure and schemas that can help to better organize huge collections. This is essential in order for these documents to be effectively and efficiently retrieved.

Dealing with unstructured information is a hot research. A growing body of work is addressing the problem of recognizing structure and schemas in documents of various types. Some areas are mainly concerned about the visual representation of documents and increasing improvements are being made in the area of pattern recognition and document layout analysis to classify documents according to structure found in their layout. On the other side, extensive research is being done in the field of machine learning to exploit attributes of documents and relationships among different documents to infer structures in large collections of documents. Important work is also being performed in the data mining and knowledge discovery community which has traditionally dealt with raw data but recently is dedicating attention to learning structure from unstructured information. In addition, Semantic Web researchers are dedicating important efforts to the problem of identifying structure and schema

in order for them to achieve ontology matching or alignment. Another related area regards the database community that has long worked with integration problems but only recently this community has started considering automatic structure and schema learning as a potential approach for schema and database integration. Finally information retrieval and extraction seek to infer structure and schemas from free text in order to build efficient information seeking models from large corpora.

The Overall Objective of the Book

The goal of this book is to present state-of-the-art methods for structure learning and schema inference. Most of the existing fields and technologies have long worked mainly in an isolated fashion even though the tasks they solve have much in common. This has led to a stall of the overall advancement to solving the problem, even though separate fields improve their performances independently on specific datasets. The automatic inference of structure is central to all approaches to organizing documents, therefore it has become important to bring together researchers from different fields and identify common challenges in order to advance the state-of-the-art in structure learning from documents. This will make possible the exploitation of methods developed in one field, from researchers of related fields who might take advantage of novelties introduced in different fields working on the same problem of learning structure in documents.

The book appreciates that an understanding of the interactions between various approaches is essential to develop synergies among different research areas in order to develop more robust methods that can attack the problem in a multi-strategic fashion. Thus the focus of this book is on:

- \* Presenting state-of-the-art approaches to learning and inferring structure and schemas from documents
- \* Assessing whether methods and techniques of one approach can be extended or exploited by the other approaches in a multi-strategic effort to the problem
- \* Assisting in the identification of new and future challenges and opportunities that will raise the bar in the state-of-the-art of several research areas attacking the same problem.
- \* Case studies and best practices from real large scale digital libraries, repositories and corpora.

#### The Target Audience

Although contributions will be open from both academia and industry practitioners and researchers, the audiences of this book are those working in or interested in joining interdisciplinary and transdisciplinary works in the areas of data mining, machine learning, pattern recognition, document analysis and understanding, semantic web, databases, artificial intelligence and digital libraries, whose main focus is that of learning structure and schemas from unstructured information. The application areas are also very broad and contributions will be open for applied works in bioinformatics, web mining, text mining, information retrieval, real-world digital libraries, data warehouses and ontology building. Specifically, audiences who are broadly involved in the domains of computer science, web technologies, applied informatics, business or management information systems are: (1) researchers or senior graduates working in academia; (2) academics, instructors and senior students!

in colleges and universities, and (3) business analysts from industries interested in data integration, information retrieval and enterprise search.

#### Topics:

Chapters should be written in a manner readable for both specialists and non-specialists. Chapters could address issues related to past, present and future theories, methods, and practices of learning structure from documents. These should be focused on next generation paradigms and with a particular focus (but not limited) to Structure Learning, Schema Integration, Schema Inference, Document Analysis and Recognition, Document Layout Analysis, Document Image Understanding, Data Mining, Data Annotation, Data Integration, Mining Unstructured Data, Learning Structure from Text, Web Mining, Text Mining, Document Databases and Digital Libraries, Database Integration, Data warehouse Integration, Ontology mapping, Ontology merging, Ontology alignment, Ontology Searching, Ontology Ranking, Ontology Evaluation, Information Retrieval, Information Extraction.

Recommended topic areas include, but are not limited to:

- \* Critical Reviews on:
  - o Theory and Strategies Fundamentals in Document Structure and Schema Learning
  - o Next Generation Technologies for Document Structure and Schema Learning
- \* Theory and Strategies Fundamentals in Document Structure and Schema Learning
  - o Developing a unifying theory of structure and schema learning from documents
  - o Document Image Analysis and Understanding, Document Analysis Systems
  - o Document Classification in Digital Libraries
  - o Ontology modeling, reuse, extraction, evolution, mapping, merging, and alignment
  - o Searching and ranking ontologies. Ontology evaluation
  - o Learning Structure and Schemas from Textual Documents for Information Retrieval and Extraction
  - o Retrieval Models and Ranking based on Structure and Schema Learning
  - o Queries and Query Analysis based on Structure Learning
  - o Recommender systems based on Structure and Schema Learning
  - o Integrating Data Warehouses Through Schema Discovery
  - o Discovering Structure and Schemas from Biological Data and Documents
  - o Merging Biological Databases Through Structure and Schema Learning
  - o Learning Structure and Schemas from Heterogeneous Domains
  - o Discovering Structure and Schemas in Social and Biological Networks
  - o Learning Structure in Graphs
  - o Web document analysis (including wikis and blogs)
  - o Communities discovery and analysis in large scale online or offline social networks
  - o Structure learning for analysis of evolution of patterns and communities in the Web
- \* Next Generation Technologies for Document Structure and Schema Learning
  - o Artificial Intelligence, Pattern Recognition and Machine Learning for Document Analysis and Understanding
  - o Symbolic, Statistical and Hybrid Learning Approaches for Structure and Schema Discovery
  - o Machine Learning for the Semantic Web
  - o Data Mining for document

- structure and schema learning
- o Distributed Computing for Learning Web Structure
- o Combinatorial Optimization in Structure and Schema Learning from Documents
- o Meta-heuristics for Structure and Schema Learning from Documents
- o Parallel Computing for Structure and Schema Learning from Documents
- o Learning structure and schemas from Stream, Spatial and Temporal Data
- o Capturing Drift in Time-Changing Documents and Data
- o High performance implementations of document analysis algorithms
- o Multi-agent systems for Document Processing and Structure Learning
- o Novel Document Representation Formalisms and Content Analysis
- o Multimedia analysis, indexing, retrieval by exploiting structure in images videos, speech/audio, etc.
- o Large scale document structure and schema learning for real-world Web scenarios
- o Large scale digital libraries by exploiting structure and schema learning
- o Large scale social network analysis based on structure and schema learning
- \* Applications and best practices in Document Structure and Schema Learning
- o Languages, Components, Programs, Knowledge Portals and/or Applications
- o User Community/Organizational Needs Response
- Developments in various settings including (but not limited to) real-world digital libraries, historical archives, document management systems, enterprise search, multimedia libraries, web search, business intelligence, economics, marketing, advertising, bioinformatics and biological networks, database and data warehouse systems, social networks, etc.
- o Performance, Scalability, Robustness, Verification, Validation, Benchmarking

#### Submission Information

Submission is possible only through invitation. Academics, researchers and practitioners are invited to submit by 20 July 2010, a 2-page manuscript proposal detailing the background, motivations and structure of their proposed chapter. Authors of accepted proposals will be notified by 1 August 2010 and will be given instructions and guidelines for chapter preparation. Full chapters are due on 30 October 2010 and should be of 8,000 words in length and/or between 25 to 30 pages long. The book is scheduled to be published in the "Studies in Computational Intelligence" book series, Springer. For information about the publisher and the book series, visit <http://www.springer.com/series/7092>. This publication is anticipated to be released in 2011.

#### Important Dates

20 July 2010:	2-page Proposal Submission Deadline
1 August 2010:	Notification of Proposal Acceptance
30 October 2010:	Full Chapter Submission (in Word or PDF)
15 December:	Notification of Full Chapter Acceptance
30 January 2011:	Revised Chapter Submission
30 February 2011:	Final Notification of Acceptance
15 March 2011:	Final Material Submission

Inquiries and submissions can be forwarded electronically (in Word or PDF) to:

Dr Marenglen Biba  
 University of New York Tirana, Albania  
 E-Mail: [marenglenbiba@unyt.edu.al](mailto:marenglenbiba@unyt.edu.al)  
 URL: <http://www.marenglenbiba.net>

or

Prof. Fatos Xhafa  
 University of London (Birkbeck), United Kingdom  
 E-Mail: [fatos@lsi.upc.edu](mailto:fatos@lsi.upc.edu)  
 URL: <http://www.lsi.upc.edu/~fatos/>

Please do not post msgs that are not relevant to the database community at large. Go to [www.cs.wisc.edu/dbworld](http://www.cs.wisc.edu/dbworld) for guidelines and posting forms.

To unsubscribe, go to <https://lists.cs.wisc.edu/mailman/listinfo/dbworld>