IQLA-GIAT Summer School in
Quantitative Analysis of Textual Data
University of Padua, 16-20 September 2013

Presentation

The analysis of textual data and large corpora is an object of research in many sectors of linguistics, computer sciences, mathematics and statistics and is used as a research tool within a number of disciplines such as psychology, sociology, sociolinguistics, education, history, political studies, communication and media studies. The recent evolution of information technologies (IT) and quantitative methods has led to a number of distinct but interrelated sectors (e.g. computational linguistics, information retrieval, natural language processing, text mining, sentiment analysis, opinion mining, etc.) with interesting industrial applications (e.g. electronic dictionaries, artificial intelligence, computer-aided translation) and, today, *text analysis* is an umbrella term referring to a wide range of qualitative, quantitative and mixed-methods approaches.

Recent studies have repeatedly stressed the need for developing, adopting and sharing interdisciplinary approaches and the IQLA-GIAT Summer School is the ideal environment for developing innovative analytical tools by pooling together the research methods of both Quantitative Linguistics and the Statistical Analysis of Textual Data.

Teaching activities will raise questions that can be answered thanks to quantitative methods implemented within a text analysis framework and other procedures which may be used to identify and compare the characteristics of texts. The aim is to discuss with students, young researchers and scholars of different disciplines the strengths, weaknesses, opportunities and threats of quantitative methods for text analysis.

Specific skills will be developed mainly with reference to the operating logic of software and tools, so that participants are trained to exploit different tools within the same environment and the same tool in different environments.

Objectives

The participants selected for the IQLA-GIAT Summer School have the opportunity to develop the necessary expertise to exploit integrated approaches to text analysis for a better understanding of the role played by theoretical issues, expected results and data and software selection.

The IQLA-GIAT Summer School is aimed at:

1. sharing information on software, corpora, relevant literature and research results;
2. promoting a dialogue among different disciplines on actual research issues;
3. developing innovative analytical tools and integrated research methods;
4. introducing students and young researchers to new strains of research and applications;
5. sharing state-of-the-art knowledge in five main areas:

- Corpora and textual data collection
- Quantitative linguistics
- Authorship attribution
- Content mapping and data visualization
- Software for computer-aided analysis of textual data

Classes

All courses are taught in English

The teaching staff includes researchers and experts from European universities: Toulouse II, ENS de Lyon (France), Trier (Germany), Athens (Greece), Neuchatel (Switzerland), Kraców (Poland), Bologna, Padova, Trieste, Roma "La Sapienza" (Italy).

Teaching activities include lectures and lab sessions, as well as workshops illustrating software and tools.

Credits

The IQLA-GIAT Summer School in *Quantitative Analysis of Textual Data* is organized by GIAT – *Interdisciplinary Text Analysis Group* (www.giat.org) in collaboration with the *International Quantitative Linguistics Association* (www.iqla.org) and the Department of *Computational Linguistics* of the University of Trier (www.uni-trier.de).

The IQLA-GIAT Summer School is a project funded by the University of Padua, coordinated by prof. Arjuna Tuzzi (University of Padua).

Schedule

The IQLA-GIAT Summer School will take place from Monday 16th to Friday 20th September 2013
The IQLA-GIAT Summer School is a full-time intensive course held on
- Monday from 10.30 am to 6.00 pm
- Tuesday through Friday from 9:30 am to 6:00 pm

Location

University of Padua
Department of Philosophy, Sociology, Education & Applied Psychology
(Dipartimento di Filosofia, Sociologia, Pedagogia e Psicologia Applicata - FISPPA)
Sociology buildings
via Cesarotti, 10/12
I-35123 Padova, ITALY

Application

The IQLA-GIAT Summer School is open to 20 participants including researchers, scholars and postgraduate students. Applicants should send a file in pdf format including:
- curriculum vitae
- personal mission statement (max 500 words)
- the request for ECTS credit (if needed - 3 CFU).

Applications should be sent to the following address: giatsummerschool.fisppa@unipd.it

Application deadline: **9th August 2013**

Selection criteria privilege foreign applicants and younger candidates to encourage participation from different countries and comply with gender equality. This will not exclude Italian candidates.
Candidates will receive a confirmation of acceptance by 16th August 2003
Tuition fee: 220 €

Subscription deadline: **5th September 2013**

Final Assessment

1) All participants will be requested to fill in a questionnaire to assess the main aspects of the IQLA-GIAT Summer School (e.g. organization, teaching, materials, facilities and equipment, expectations, satisfaction rate, suggestions etc.).
2) If they wish to receive ECTS credits (3 CFU), participants shall take a written test including 30 multiple choice questions (one correct answer out of four). The passing grade is 15 correct answers. Grades will reflect the sum of all correct answers (one point) according to the following range: A (29-30 points); B (25-28 points); C (21-24 points); D (17-20 points); E (15-16 points).

Secretariat

Claudia Andreatta, Ph.D.

giatsummerschool.fisppa@unipd.it

For any further information and details about terms, deadlines, application forms and payment modalities, please contact: giatsummerschool.fisppa@unipd.it or arjuna.tuzzi@unipd.it.

Provisional program

General topics
1. Building Corpora: caveats, opportunities and examples (Stefano Ondelli, University of Trieste, Italy)
2. Collecting corpora on the Web: the dubious joys of scraping web pages (Alberto Cammozzo, IT expert, Italy)
3. Qualitative and Quantitative Approaches (Alberta Contarello, University of Padova, Italy)
4. Methodology: measuring and testing (Reinhard Köhler, University of Trier, Germany)

Quantitative Linguistics
5. Measures of Lexical Richness (Sven Neumann, University of Trier, Germany)
6. Linguistics Laws (Reinhard Köhler, University of Trier, Germany)
7. Quantitative Syntax Analysis (Sven Neumann, University of Trier, Germany)

Authorship Attribution
8. Distance-Based Models in Authorship Attribution (Jacques Savoy, University of Neuchatel, Switzerland)
9. Machine Learning methods in Authorship Attribution and Profiling (George Mikros, University of Athens, Greece)
10. Mathematical Models of Textual data with applications to Authorship Attribution (Mirko Degli Esposti, Alma Mater University of Bologna, Italy)

Software (Workshop) and other Lab-Sessions
11. LAB – Introduction to Iramuteq (Pierre Ratinaud, University of Toulouse II, France)
12. LAB – Introduction to TXM (Serge Heiden, ENS de Lyon, France)
13. LAB – Introduction to TaLTac2 – Lexical and Textual Automatic Processing for Corpus and Content Analysis (Isabella Chiari, University of Rome "La Sapienza", Italy)
14. LAB – Introduction to R package "Stylo" (Maciej Eder, Kraców, Poland)
15. LAB – Introduction to Tree Tagger (Teresa Musacchio, University of Padova, Italy)
16. LAB – Probability distributions (Andrej Beliankou, University of Trier, Germany)
17. LAB – Lab session on profiling (George Mikros, University of Athens, Greece)