ESF Exploratory Workshop on

COMPUTING IN HUMANITIES

Sofia (Bulgaria) 8-9 April 2015

SCIENTIFIC REPORT

Convened by:
Svetla Koeva, Tsvetana Dimitrova
and Marko Tadic

1. Executive summary

The workshop *Computing in Humanities* was organized by the Institute for Bulgarian Language "Prof. Lyubomir Andreychin" at the Bulgarian Academy of Sciences as a local host within the framework and with the support of *NeDiMAH: a Network for Digital Methods in the Arts and Humanities*. The two-day workshop was held on April 8-9, 2015, in *Mercan Hall, Ramada Hotel, Sofia* (see the workshop programme in section 4 of the Scientific report). The workshop aimed to support interdisciplinary work between research communities in the humanities, with a special focus on digital methods in the humanities research. The main objectives of the workshop were to encourage the interdisciplinary work through a collaborative research; to focus priority research areas; to create links between research communities in humanities; and to provide opportunities for dissemination of the research. Early-stage researchers in humanities were given the opportunity to participate into the efforts for developing an interdisciplinary research community that would further study, develop, refine and share various digital methods. The workshop was in line with the aims of the *NeDiMAH Network* towards promoting and application of advanced Information and Communication technologies in the humanities across Europe.

The worksop was attended by 26 participants and 3 lecturers, most of whom were early stage researchers (average participation of 25 persons per session). The workshop brought together experts and students from 6 countries: Denmark, Luxembourg, Croatia, Hungary, Romania, and Bulgaria. 23 of the participants were from Bulgarian research institutions and universities: the Institute for Bulgarian Language (IBL-BAS) (four early stage researchers and three Ph.D. students in computational linguistics, terminology, and dialect studies); the Institute for Literature (IL-BAS) – two Ph.D. Students (working on comparative literature, contemporary Bulgarian literature and history of Bulgarian literature); the Institute of Ethnology and Folklore Studies with Ethnological Museum (IEFEM-BAS) (an assistant professor in Balkan ethnology working on the topics of migration and integration of Bulgarian Muslim population and Bulgarian Turks); Cyrillo-Methodian Research Centre (CMRC-BAS) (a young researcher working on South Slavonic mediaeval manuscripts and biblical studies); the Institute for Art Studies (IAS-BAS) (a post-doc working on virtual archaeology, architectural graphics, digital archaeology and architecture). There were attendees from Sofia University – two Ph.D. students (both pursue their Ph.D. in historical linguistics with a focus on comparative methods and digital humanities) and two Master's students (one in Computational Linguistics at the Faculty of Slavic Studies, and the other – in Archaeology, with a focus on digital and quantitative methods), five assistant professors from Plovdiv University (with research interests in comparative and computational methods in linguistics, cognitive linguistics and translational studies); and an assistant professor at the Department of Language Studies at the University of Food Technologies (UFT) in Plovdiv whose research is focused on formal linguistics, argument structure and lexical semantics. There were two young researches from the Institute of Linguistics at the Hungarian Academy of Sciences (NYTUD-MTA), and one Ph.D. student from the Institute for Artificial Intelligence at the Romanian Academy (RACAI). The accommodation and dinner for 6 participants from abroad and 6 participants from Plovdiv were provided by the Ramada Hotel. Four coffee breaks and two lunches for all participants were organized at the same place. At the end of the workshop, the participants were given certificates for attending it and t-shirts as a symbol that they all belong to the community of digital humanities researchers.

Lectures and presentations were given across the two days and there was ample time for discussions at the end of each of the four sessions. The workshop was focused on three topics

(*Transcribing and Describing Primary Sources in TEI XML; Extraction and Visualization of Historical Sources (From Text Interpretation to Data to Networks); Text Technologies for Humanities Research),* and for each topic there was a combination between an initial overview of the available resources, research networks and data repositories, some highlighted examples of emerging technologies, hands-on sessions, and open group discussions.

Two specific case studies were presented in detail: Old and Early-Modern Icelandic sources and English early sources (with both texts and images) and historical sources, chosen with the aim of covering a broad variety of document types. Different applications were used to elaborate TEI XML source description and network data extraction from historical sources using qualitative data analysis techniques: Oxygen, NodeXL, and Palladio. The presentations, handouts, and links to different materials and applications on the internet were circulated among the participants.

The main results of the workshop can be summarized as follows:

It was shown that the digital technologies have the power to transform humanities research, making it easier and more efficient, answering the research questions more systematically, enabling new ways of working and collaborating, opening up new challenges and creating new research paradigms.

It was demonstrated that the standard vocabulary is essential for the retrieval, sharing, and use of texts. Such vocabulary is the Text Encoding Initiative (TEI) XML format for text, since it provides a standard way of representation for dates, language symbols, special symbols, images, metadata description (title, author, year of publication, etc.), and so on. The mark-up languages are also useful to make use of the options for adding explanatory notes and comments to texts but in general, they can be used to mark and retrieve any particular element of the text.

The challenge to systematizing text interpretation was discussed: a formal category system should represent the different meanings in as much detail as necessary for a particular purpose. It was demonstrated how the context knowledge analysis of the text can help to identify what it stands for in any given case. The discussion was focused on data extraction from unstructured text and a way to visualize it was shown. It was shown that the networks created from pre-existing data sets need to be considered within the context in which they were created. On the other hand, the networks created from unstructured text pose challenges on top of this: interpretations are highly individual and depend on viewpoints and context knowledge.

The *Computing in Humanities* workshop enjoys very positive feedback from the participants with the observation that the humanities research community is in need of events to support and strengthen the links in the community (and between research communities in different areas in the humanities).

2. Scientific content of the event

Prof. Matthew James Driscoll – a senior lecturer in Old Norse philology at the Arnamagnæan Institute, a research institute within the Faculty of Humanities at the University of Copenhagen, presented the TEI approach and the TEI-conformant XML encoding for description of (textual and visual) sources in the humanities. His two-part lecture *Transcribing and Describing Primary Sources in TEI XML* concluded with a hands-on session on TEI XML source description using Oxygen (installed beforehand by the participants following instructions provided by Prof. Driscoll). The samples given were from Old and Early-Modern Icelandic sources and English early sources (with both texts and images); the

applicability of the TEI scheme for the sources that some of the participants' worked on was also discussed. During the lectures and the practical exercises the students found answers of the following questions: how much of the information in an original source, be it a manuscript, charter or early printed book, should be included in a transcription or edition; is the distinction between the 'substantives', the actual words of the text, and the accidentals', features such as spelling, punctuation, page layout etc., a useful one; are the 'accidentals' really of no interest or value, etc. In that way, in addition to covering the fundamentals of transcription and description using TEI, the workshop exposed participants to methods by which the encoded text can be presented and/or published electronically.

The lectures of Dr. Marten During on *Extraction and Visualization of Historical Sources (From Text Interpretation to Data to Networks)* were held as two afternoon sessions on both days with extensive and very interesting and useful hands-on sessions (the work programme, presentations and materials were also made available on the internet and installation instructions were provided beforehand by Dr. During). Dr. During is a researcher in the Digital Humanities Lab at the European Virtual Knowledge Center in Luxembourg. He works with Social Network Analysis and is interested in the applicability of software and the Internet for historical research, teaching and public outreach. The lectures show how network visualizations can help humanities scholars reveal hidden and complex patterns and structures in textual sources. The lectures explained how to extract network data (people, institutions, places, etc.) from historical sources through the use of non-technical methods developed in Qualitative Data Analysis (QDA) and Social Network Analysis (SNA), and how to visualize this data with the platform-independent and particularly easy-to-use application Palladio. During the hands-on sessions, the participants worked on network data extraction from historical sources using qualitative data analysis techniques (and applications NodeXL and Palladio, with a brief overview of their functionalities). The participants were divided into five groups (working in two different conference halls) to extract and describe/encode data from a first-person narrative of a Jewish survivor of the Holocaust using an existing coding scheme. The afternoon session on the second day was primarily dedicated to practical work on data organisation and description using data visualisation software for visualising the relations between entities/persons mentioned in the historical narrative and described during the hands-on session on the previous day.

The morning session on the second day of the workshop was dedicated to problems of *Text Technologies for Humanities Research* presented by Prof. Marko Tadic from the Department of Linguistics of the Faculty of Humanities and Social Sciences at the University of Zagreb. Prof. Tadic is also an associated member of the Croatian Academy of Sciences and Arts and a member of the Standing Committee for the Humanities of the European Science Foundation (2009-2012). He is a (co)author of important language resources for the Croatian language such as the Croatian National Corpus, Croatian Dependency Treebank and the portal Language Technologies for Croatian Language. His two-part lecture involved an overview of the state-of-the-art in text technologies for the humanities, and a discussion on the problems and opportunities for humanities research and building links to access and use the knowledge in the humanities created by the development of language and text technologies. Prof. Tadic gave an overview of the available resources, research networks and data repositories (such as CLARIN, META-NET, links with DARIAH for the image and visual documents and data), and the XLike technology for cross-lingual knowledge extraction (available for English, German, Spanish, Chinese, Hindi, Catalan, and Slovenian). Prof. Tadic presented an extensive discussion on the applications and uses of the language and text technologies in other research fields in the humanities (such as archaeology, anthropology, ethnology, art studies, visual and image studies, among others).

3. Assessment of the results, contribution to the future direction of the field, outcome

In general, researchers in the humanities view communication and information technologies both as a challenge and as a big opportunity. The main question is how the technologies are being communicated to the humanities researchers. It has to be acknowledged that the scholars in these fields may not always use the most appropriate techniques and methods in analyzing, retrieving and sharing data. In this regard, the workshop was extremely well received and we can conclude the the main objectives of the workshop: to encourage interdisciplinary work through collaborative research; to focus priority research areas; to create links between research communities in the humanities; and to provide dissemination of research were successfully achieved. Moreover, the early-stage researchers saw a large number of potential developments based on the workshop presentations and discussions. They saw the value of the shared knowledge and integrated and accessible data collections and methods for encoding, description and visualisation of the data they all work with (such as language data, incl. dialectal data; manuscript data; information on migrants and their integration and re-integration; term extraction, etc.). After the workshop, the participants expressed their willingness to continue to use the applications provided for their own research purposes and also expressed interest in knowing research groups that work with the same methods (for example, for digitisation of manuscript descriptions).

The follow-up actions that we are determined to undertake at this point are the following:
- distribution of the workshop materials (lecture presentation and materials from the hands-on session);
- creation of an interdisciplinary network of young people (including those researcher who were not able to attend the workshop);
- support for the links between the researchers, the lecturers and research groups they work in;
- dissemination of information about the workshop and research opportunities it can create.

4. Final programme

FINAL PROGRAMME

Date (Wednesday, 8th April) 2015

09.00-09.15	*Registration*
09.15-09.30	Welcome
	Svetla Koeva (Institute for Bulgaria Language, Sofia, Bulgaria)
09.30-13:00	**Morning Session**
09.30-10:30	Lecture 1 Transcribing and Describing Primary Sources in TEI XML
	Matthew James DRISCOLL (University of Copenhagen, Denmark)
10.30-11.00	*Coffee / Tea Break*
11.00-12.00	Lecture 2 Transcribing and Describing Primary Sources in TEI XML
	Matthew James DRISCOLL (University of Copenhagen, Denmark)

12.00-13.00	**Discussion**
13.00-14.30	*Lunch*
14.00-17:30	**Afternoon Session**
14.00-15.30	Lecture **3** Data Extraction and Visualization of Historical Sources
	Marten DÜRING (Digital Humanities Lab, Luxembourg)
16.30-16.00	*Coffee / tea break*
16.00-17.30	Lecture **4** Data Extraction and Visualization of Historical Sources
	Marten DÜRING (Digital Humanities Lab, Luxembourg)

Date (Thursday, 9th April) 2015

09.00-13:00	**Morning Session**
09.00-10.30	Lecture 1 Text technologies for Humanities Research
	Marko TADIĆ (University of Zagreb, Zagreb, Croatia)
10.30-11.00	*Coffee / Tea Break*
11.00-12.00	Lecture 2 Text technologies for Humanities Research
	Marko TADIĆ (University of Zagreb, Zagreb, Croatia
12.00-13.00	**Discussion**
12.30-14.00	*Lunch*
14.00-17:00	**Afternoon Session**
14.00-15.30	Lecture **3** Data Extraction and Visualization of Historical Sources
	Marten DÜRING (Digital Humanities Lab, Luxembourg)
15.30-16.00	*Coffee / tea break*
16.00-17.00	**Discussion**
17.00	*End of Workshop*

5. Final list of participants
(name and affiliation is sufficient; the detailed list should be updated on-line directly)

Convenor:
1. Svetla KOEVA
Department of Computational Linguistics
Institute for Bulgarian Language
Bulgarian Academy of Sciences
52 Shipchenski prohod Blvd., Bldg. 17
1113 Sofia
Bulgaria
svetla@dcl.bas.bg

Co-Convenor:
2. Tsvetana DIMITROVA
Department of Computational Linguistics
Institute for Bulgarian Language
Bulgarian Academy of Sciences
52 Shipchenski prohod Blvd., Bldg. 17
1113 Sofia
Bulgaria
cvetana@dcl.bas.bg

3. Marko TADIĆ
Institute of Linguistics
Faculty of Philosophy
University of Zagreb
Ivana Lučića 3
10000 Zagreb
Croatia
marko.tadic@ffzg.hr

Participants:
4. Matthew James DRISCOLL
Department of Nordic Research
University of Copenhagen
Njalsgade 136
DK-2300 Copenhagen S
Denmark
mjd@hum.ku.dk

5. Marten DÜRING
Digital Humanities Lab
Centre virtuel de la connaissance sur l'Europe
Château de Sanem
L-4992 Sanem
Luxembourg
marten.duering@cvce.eu

6. Zsófia LUDÁNYI
Department of Language Technology and Sociolinguistics
Research Institute for Linguistics
Hungarian Academy of Sciences
1394 Budapest, P.O. Box 360
Hungary
ludanyi.zsofia@nytud.mta.hu

7. Réka KOVÁCS
Department of Language Technology and Sociolinguistics
Research Institute for Linguistics
Hungarian Academy of Sciences
1394 Budapest, P.O. Box 360

Hungary
kovacs.reka@nytud.mta.hu

8. Eric CUREA
Research Institute for Artificial Intelligence
"Mihai Drăgănescu" (RACAI)
Romanian Academy
Calea 13 Septembrie, No. 13
Casa Academiei
050711 Bucharest
Romania
eric@racai.ro

9. Mila MAEVA
Institute of Ethnology and Folklore Studies with Ethnographic Museum
Bulgarian Academy of Sciences
6 Moskovska St.
1000 Sofia
Bulgaria
mila_maeva@mail.bg

10. Iva TRIFONOVA
Cyrillo-Methodian Research Centre (CMRC)
Bulgarian Academy of Sciences
13 Moskovska St.
1000 Sofia
Bulgaria
iva_3@abv.bg

11. Stela TASHEVA
Institute of Art Studies
Bulgarian Academy of Sciences
Acad. G. Bonchev St., bldg. 16
1113 Sofia
Bulgaria
stelabt@gmail.com

12. Elena BORISOVA
Institute for Literature
Bulgarian Academy of Sciences
52 Shipchenski prohod Blvd., Bldg. 17
1113 Sofia
Bulgaria
elena39789@abv.bg

13. Vassil VASSILEV
Institute for Literature
Bulgarian Academy of Sciences
52 Shipchenski prohod Blvd., Bldg. 17

1113 Sofia
Bulgaria
vivvox9@gmail.com

14. Svetlozara LESEVA
Institute for Bulgarian Language
Bulgarian Academy of Sciences
52 Shipchenski prohod Blvd., Bldg. 17
1113 Sofia
Bulgaria
zarka@dcl.bas.bg

15. Maria TODOROVA
Institute for Bulgarian Language
Bulgarian Academy of Sciences
52 Shipchenski prohod Blvd., Bldg. 17
1113 Sofia
Bulgaria
maria@dcl.bas.bg

16. Atanaska ATANASOVA
Institute for Bulgarian Language
Bulgarian Academy of Sciences
52 Shipchenski prohod Blvd., Bldg. 17
1113 Sofia
Bulgaria
nassi_n@abv.bg

17. Nikolay ZHELYAZKOV
University of Food Technologies
26 Maritsa Blvd.
4003 Plovdiv
Bulgaria
nikolay.zhelyazkov@gmail.com

18. Kremena DULGEROVA
Institute for Bulgarian Language
Bulgarian Academy of Sciences
52 Shipchenski prohod Blvd., Bldg. 17
1113 Sofia
Bulgaria
emito_s@abv.bg

19. Verginiya D. GEORGIEVA
Institute for Bulgarian Language
Bulgarian Academy of Sciences
52 Shipchenski prohod Blvd., Bldg. 17
1113 Sofia
Bulgaria

vdamgeorg@abv.bg

20. Valentina STEFANOVA
Faculty of Slavic Studies
Sofia University
15 Tsar Osvoboditel Blvd.
1504 Sofia
Bulgaria
valentina.stefanova87@gmail.com

21. Mariyana NIKOLOVA
Faculty of Slavic Studies
Sofia University
15 Tsar Osvoboditel Blvd.
1504 Sofia
Bulgaria
m.staykova@abv.bg

22. Viktoriya PETROVA
Faculty of Slavic Studies
Sofia University
15 Tsar Osvoboditel Blvd.
1504 Sofia
Bulgaria
v.k.petrova@abv.bg

23. Adelina RADEVA
Faculty of Slavic Studies
Sofia University
15 Tsar Osvoboditel Blvd.
1504 Sofia
Bulgaria
radeva.adelina@gmail.com

24. Rositsa DEKOVA
Department of English Studies
Plovdiv University
24 Tsar Assen St.
4000 Plovdiv
Bulgaria
rosdek@gmail.com

25. Zlatka CHERVENKOVA
Department of English Studies
Plovdiv University
24 Tsar Assen St.
4000 Plovdiv
Bulgaria
zlatche@yahoo.com

26. Magdalena GOGALCHEVA
Department of English Studies
Plovdiv University
24 Tsar Assen St.
4000 Plovdiv
Bulgaria
mag. gogalcheva@gmail.com

27. Magdalena GOGALCHEVA
Department of English Studies
Plovdiv University
24 Tsar Assen St.
4000 Plovdiv
Bulgaria
polly.petcova@gmail.com

28. Vesselina KOYNAKOVA
Department of English Studies
Plovdiv University
24 Tsar Assen St.
4000 Plovdiv
Bulgaria
vescoy@gmail.com

29. Petya NESTOROVA
Department of Language Teaching
University of Food Technologies
26 Maritsa Blvd., bldg. 2, fl. 2
4000 Plovdiv
Bulgaria
petyanes@gmail.com

6. 6. Statistical information on participants

6.6.1. Summarised information

Age range	25-40:	40-43:		expert	expert	expert
No participants	**22**	**3**				
Countries of origin	BG	HU	RO	HR	DK	LU
No participants	**22 + 1 covenor**	**2**	**1**	**1**	**1**	**1**
M/F repartition	M	F				

No participants	3+3 lecturers	22 + 1 convenor				
Disciplinary fields	*Language studies*	*Literary studies*	*Historical studies, ethnology*	*Art studies*		
No participants	**18**	**3**	**3**	**1**		
Institutions	*Research institute*	*University*				
No participants	**15**	**10**				
Status (career stage)	*Assistant professor, PhD*	*PhD student*	*MA student*			
No participants	**15**	**8**	**2**			

6.6.2. Disciplinary fields represented (excluding lecturers and convenor)

Ethnology, Balkan ethnology: 1
Historical studies, archaeology: 1
Historical studies, manuscript and language studies: 1
Art Studies, architecture: 1
Language studies, computational linguistics: 6
Language studies, linguistics: 3
Language studies, historical linguistics: 3
Language studies, dialectology: 1
Language studies, lexicology: 2
Language studies, cognitive linguistics: 3
Literary studies: 3