The MATE Workbench

Laila Dybkjær and Niels Ole Bernsen

Natural Interactive Systems Laboratory, University of Southern Denmark Science Park 10, 5230 Odense M, Denmark laila@nis.sdu.dk, nob@nis.sdu.dk

Thematic area: language corpora and related tools.

Keywords: spoken dialogue corpora, multilevel annotation, information extraction, workbench.

1. Abstract of technical content to be demonstrated

The aim of the European Telematics project MATE (Multilevel Annotation Tools Engineering) has been to facilitate the use and reuse of spoken language resources by addressing theoretical issues and implementing practical solutions. MATE was launched in 1998 in response to the increasing need - not least in the area of spoken language dialogue systems research and development - for tools and standardisation efforts in support of efficient markup of spoken dialogue corpora at multiple levels.

To provide a solid basis for the coding standard to be proposed by MATE, more than sixty existing coding schemes belonging to five different coding levels (i.e. prosody, (morpho-) syntax, co-reference, dialogue acts, and communication problems) and their cross-level interaction were reviewed (Klein et al. 1998). On this background, the MATE markup framework (Dybkjær et al. 1998) was proposed as a standard which can facilitate uniform description of schemes across levels. For each level, one or more of the reviewed coding schemes were taken as the starting point for the definition, following the MATE markup framework, of one or more best practice coding schemes (Mengel et al. 2000).

The MATE software workbench (Isard et al. 1998, Isard et al. 2000) supports the MATE markup framework and incorporates the MATE best practice coding schemes. It is implemented in Java 1.2. It has been tested on Unix (Solaris) and Windows (NT and 98) but should run on any platform for which Java 1.2 is available. The workbench has a modular architecture which facilitates the addition of new modules and new tool functionality by its users. XML is used for internal file representation.

The MATE workbench has the following major components: an internal database which is an in-memory representation of a set of hyperlinked XML documents; a query language and processor which are used to select parts of this database for subsequent display or processing; a stylesheet language and processor which respectively define and implement a language for describing structural transformations on the database; a display processor which handles the display and editing actions; and a user interface which handles file manipulation and tool invocation.

Stylesheets are used for specifying the visual presentation of data to users. Stylesheets are written in the MATE Stylesheet Language (MSL). The emerging standard in this area is XSLT, but XSLT was not fully defined when the workbench was being designed, and lacked various necessary functionalities. It was therefore decided to implement MSL which uses the MATE query language but is otherwise similar to XSLT.

The MATE workbench offers support for annotation of spoken dialogue corpora and information extraction from annotated corpora as its two main functionalities. Both aspects will be demonstrated in detail during the presentation. The annotation support includes:

• Ready-to-use best practice coding schemes for the levels mentioned above. Stylesheets are included which allow the user to view and annotate files in an appropriate way using any of these best practice schemes. Example dialogues are also provided which illustrate how

a tagged dialogue and the accompanying tag set will be shown when a given best practice scheme is being applied.

- Easy addition of new coding schemes for existing or new levels. New coding schemes are entered by using the coding module editor. A coding module is the standard type of coding scheme description proposed by MATE. A coding module prescribes what constitutes a coding, including markup representation and relations to other codings.
- Editing of files. MATE operates with three basic kinds of files. XML files are coding files. MSL files are stylesheets used for visualisation of codings and for annotation support. Run files specify which stylesheet to apply to which XML file. Any XML, MSL or run file in a MATE project folder can be opened and edited.
- Creation of new project folders and addition or creation of new files, including header file documentation.
- An audio tool for listening to speech files and displaying the sound file as a waveform.
- Conversion to XML format of files in other formats. For the moment, conversion from XLabels and BAS Partitur to XML is enabled but new converters can easily be added, including converters which export from XML to other formats, such as HTML.

As regards information extraction, a query tool is available for selecting the document(s) to be searched and for specifying the information to look for. The information which can be extracted includes statistical information.

An online help facility is available and may be consulted at any time during use of the MATE workbench..

Acknowledgements

We gratefully acknowledge the support to the MATE project provided by the European Commission's Telematics/Language Engineering Programme. We would also like to thank all MATE partners. Without the very considerable joint efforts of the project consortium it would not have been possible to build the MATE Workbench.

References

MATE deliverables are available from the MATE web site at http://mate.nis.sdu.dk.

- Dybkjær, L., Bernsen, N. O., Dybkjær, H., McKelvie, D. and Mengel, A.: The MATE Markup Framework. MATE Deliverable D1.2, 1998.
- Isard, A., McKelvie, D., Cappelli, B., Dybkjar, L., Evert, S., Fitschen, A., Heid, U., Kipp, M., Klein, M., Mengel, A., Møler, M. B. and Reithinger, N.: Specification of Workbench Architecture. MATE Deliverable D3.1, 1998.
- Isard, A., McKelvie, D., Mengel, A., Møler, M. B., Grosse, M. and Olsen, M. V.: Data Structures and APIs for the MATE Workbench. MATE Deliverable D3.2, 2000.
- Klein, M., Bernsen, N. O., Davies, S., Dybkjær, L., Garrido, J., Kasch, H., Mengel, A., Pirrelli, V., Poesio, M., Quazza, S. and Soria, C.: Supported Coding Schemes. MATE Deliverable D1.1, 1998.
- Mengel, A., Dybkjær, L., Garrido, J., Heid, U., Klein, M., Pirrelli, V., Poesio, M., Quazza, S., Schiffrin, A. and Soria, C.: MATE Dialogue Annotation Guidelines. MATE Deliverable D2.1, 2000.

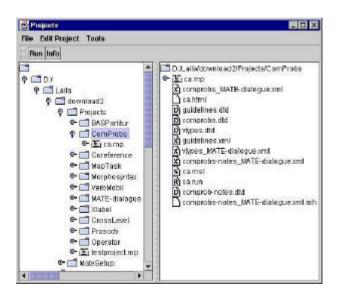
2. Presentation script

The MATE workbench can be downloaded from http://mate.nis.sdu.dk.

Starting the MATE workbench

Upon start-up the following two windows will appear (the control window and the projects window).





Annotation

Annotation can be done using one of the MATE best practice coding schemes or users may add their own coding schemes. The two windows below show annotation based on the MATE MapTask scheme and the MATE communication problems scheme, respectively.

File Edit Tools	Help
Instruct	SPLAY giver
explain	okay
check	PLAY giver instruct
query-yn	starting off we are above a caravan park PLAY follower
query-w	mmhmm
align	PLAY giver
reply-y	we are going to go due south straight south and then we're going to g turn straight back round and
reply-n	head north past an old mill on the right hand side
reply-w	due south and then back up again
acknowledge	PLAY giver
ready	yeah
clarify	PLAY giver
uncodable	south and then straight back up again with an old mill on the right and you'te going to pass on the left-hand side of the mill
	PLAY follower right okay
)

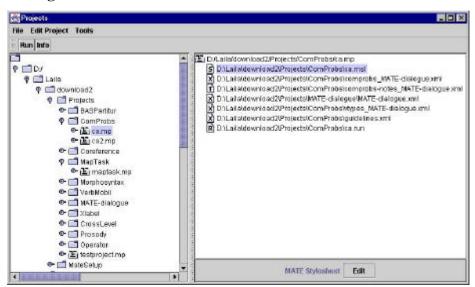
₿C	ommunication Problems				
File	Edit Tools Help				
Cami	munication Problems: comprobs_MATE-dialo	gue	1 14	000	we write.
5 U 5 U	weitrome to hips 97 version 3 point 2 fm ready to start hallo N show me a map of pacifica	904-1 905-1 907-1	N1	009 0010 564 565 6611 566 567 6612	Be orderly. Highlight asymmetries. State your capabilities State how to interact. Be aware of user background knowledge. Be aware of user inferences. Adapt to novices and experts. Be aware of user expectations.
5	ok clear where are the people		N2	508 0013 509	Cover the domain. Enable mata-communication. Enable system repair.
N1	Add New Note We do not have the graphics part of the may be instructions and feedback on 1 not know about. Noneover, we don't im intended users have access to. The m problems is made exclusively on the b of the claipgue. Probolity a mail is displayed on the so	he screen whi ow for sure whi thonal informa arkup of comm asis of the sp	ich we do ho the ation nunicatic seech par	862-2 864-1 505-1	etc.). The utterance should be reflected back at the user in the response to reassure them that they are being correctly understood. It is usually not a good idea to use coreferences in teedback situations because it does not become clear exactly what the system is talking about. There is apparently no communication of what the system can and cannot do. No instructors are provided to the user on how to interact with the system.
NJ	necessary feedback. If this is the case 802. It is not clear which of the two trucks th	there is no vi	olation of	9G7-1	The user is assumed to know how to use the system - there is no offer of more datailed information for the user should this n be the case.

Listening to audio files

Audio files can be selected in the projects window and played. The window below will then appear.



Editing a file



MATE XML, MSL (stylesheets), and run files in MATE project folders can be edited by selecting the file and clicking Edit in the projects window (see window above).





Basic Coding/Andor Title = "Communic Size = (800,600); Siylesheet = "ra.m MATEFile = "rama	ation Probl sF;	ems"; =dialoque.con";	
1	50		
1	- E6		

Extracting information from annotated corpora

Query History Results 5	tastics	
(\$x comprote)	(By verial)	+
	(\$ viref +) and	
(5 X (comprobe) • 5 X (comprobe) • 1	A11
() x (comprofis)	▼ d	Tald
		") Add

From the query window one or more MATE documents can be selected and queried (se window above). Results are shown as exemplified below.

🖉 Query Win	daw.			
File				
Decuments	Query History	Results	Statistics	
			(Sy viref) 🗢	
qutup_1		- 1		-
quitup_2	€y.	3		
qutup_3	5Y.	_	<pre>cvbrefid="VT2" href="vbypes_MATE-dialogue.xmWid(SG5-1)" > </pre>	
utup_4	By:	-	<pre>cvtrefid="VT3" href="vtypes_MATE-dialogue.xm#id(867-1)" > </pre>	
utup_5	\$γ.		<pre>cvtrefid="VT4" href="vtypes_MATE-dialogue.xmWid(SG2-1)" > </pre>	
utup_6	₿y.		ovtref id="VT5" href="vtypes_MATE-claigue.xm#id(36(2-1)" >	
htup_7	€y.		<pre>cvtrefid="VT6" href="vtypes_MATE-clalogue.xm#d(SG2-2)" > <atref></atref></pre>	
utup_8	\$y.		<pre>cvtrefid="VT7" href="vtypes_MATE-dialogue.xmWid(SG2-2)" > </pre>	
utup_9	\$y:		ovtref id="VT8" href="vtypes_NATE-dialogue.xm#id(BG2-1)" > <atref></atref>	
utup_10	Sy.		<pre>cvtrefid="VT9" href="vtypes_MATE-dialogue.xmWid(SG2-1)" > </pre>	
utup_11	By:		ovtref id="vT10" href="vtypes_MATE-dialogue.cm#id(562-1)" >	
jutup_12	₽y.	-	<pre>cvtrefid="\/T11" href="vtypes_MATE-dialogue.cm##d(SG2-1)" ></pre>	
1	Sy.		<pre>cvbrefid="vT12" href="vtypes_MATE-dialogue.com#kd(SG2-1)" > </pre>	

Adding a new coding module

Users may add new coding modules (coding schemes) themselves via the coding module editor shown here:

🔹 Mate Coding Module Editor File Edit Functions	- New coding module	
Tree Coding module XML D		
Name	Attribute name	
Coding purpose	FirstAttr	¥
Data sources	Туре	
Module references Warkup declaration	ENUM	+
Entities	Values	
Elements MyElement	Values	
- Id	yes no	
EinstAttr Description Example		
Coding procedure Creation notes		
	Presence	-1
	#DEFAULT	
	Example	
1.00		
	m dén de la complete	

Adding a new project

A new MATE project folder can be created as follows:

MAI	E query
9	Please enter name of new project to crea
6	D:/Laila/download2/Projects/ComProbs/
	OK Cancel

Adding a new file

Existing files may be added to a project folder or new files may be created.

Add files to	project			5
Look in:	ComProbs	•		
_ cazmp		12014 10	-1 0 C	Ъ
comprob-ni	otes.dtd			1
comprobs-r	notes_MATE-dialogue.xml			
comprobs-r	notes_MATE-dialogue.xml.mh			
comprobs				
comprobs	MATE-dialogue.cml			
guidelines.	dtd			
] guidelines:	ani			
File name:	guidelines xml			Open
Files of type:	All Files ('.')		•	Cancel

New	file
Ĵ	Choose file format
	MATE XML File (.xml) 🗸
	MATE Run File (.run) Apply MATE Stylesheet to MATE XML File ()
	MATE XML File (.xml)
	MATE Stylesheet (.msl)

	new files are put			
	d2\Projects\ComProbs	í		
File name for the	e new MATE XML File-			
Default MATE Sty	Jochoot			
Derduit MATE SU	plesheet	2	21	
			>>	×
Files used/referr	ed to by this file			(C.)
			>>	×
Use links to files	or copy them to proje	ct directory		
Jse links to files				
				_
Coded by (name	and profession)			
Coded by (name	and profession)			
Coded by (name Date of coding—	and profession)			
Date of coding—	and profession)			
	and profession)			
Date of coding— Version————		1		
Date of coding— Version————	and profession)			
Date of coding— Version———— DTD file used or	Coding module applied		>>	>
Date of coding— Version———— DTD file used or			nana. A	>
Date of coding— Version———— DTD file used or The project cont	Coding module applied ext in which the codin		nana. A	
Date of coding— Version———— DTD file used or	Coding module applied ext in which the codin		nana. A	3
Date of coding— Version— DTD file used or The project cont Purpose of codir	Coding module applied ext in which the codin	g was made	e	
Date of coding— Version— DTD file used or The project cont Purpose of codir	Coding module applied ext in which the codin ng	g was made	e	>
Date of coding Version DTD file used or The project cont Purpose of codin Contact details f	Coding module applied ext in which the codin ng	g was made	e	>

As a minimum, the new file must be given a name, and an applicable DTD or coding module must be specified. However, it is recommended that additional header information is provided as well.

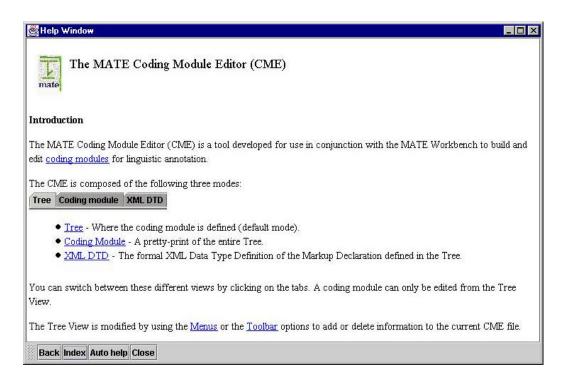
Import from, and export to, other file formats

Files may be imported from other formats to XML. Below is shown import from BAS Partitur.

MATE Conversion	m tools	
Choose conversion	tool:	
Bas Partitur >>> XI	AL.	,
Files selected for co	Inversion:	
D.Cana Coverioa 024	Frageria da April	arliturie032ach1_001_RGM.par

Getting help

A help window can be opened which provides information on various workbench issues. Below is shown the topmost help page for the coding module editor.



3. Hardware and software requirements

We will bring a portable PC with the MATE workbench installed.

However, we would very much like the local organisers to provide us with either a data projector or a large screen. If the demonstration is a, say, 20 minute presentation in a lecture hall in front of a large audience, a data projector is needed. If the demonstration is on exhibition- or poster- like conditions where people come and go and ask questions and discuss, we would prefer a large screen.