



Universiteit Utrecht



UNIVERSITEIT VAN AMSTERDAM



Demonstration scenario

M2 QuaMeRDES (CLARIN-NL-12-005)

Jasmijn Van Gorp

1. Executive summary

This milestone (M2) describes the demonstration scenario for QuaMeRDES. In it we detail how a media researcher uses the tool for her own research in order to crystallize the features and functionalities of the new QuaMeRDES-tool. The material is organized in five steps around the research practice of the indicative persona 'Corine', a Dutch 42-year-old senior researcher in Media Studies with experience in using digital television and newspaper archives. First, we set the scene by stating the main goals of Corine's research (2.1.). Every step of the description of the demonstration scenario (except 2.6.) starts with one or more research questions. We then follow with a description of her actions before discussing the hypothetical findings of her actions.

To arrive at this scenario, we proceeded as follows. Previous research pointed out that media researchers first explore, then contextualize / analyze, and in a last phase write up their research.¹ We implemented the three research phases of the media researcher in the scenario: exploration is done in the first step (2.2.); contextualization / analyzing in the three following steps (2.3., 2.4. and 2.5); and writing up the research is done in the last step (2.6). In addition, media researchers often revise their research questions while collecting data in the exploration and contextualization phase.² We echoed the revising of the research questions by starting with a broad topic (Russia) and ending up with a specific research question (representation of cultural identity in programmes about/in Russia). For contextualizing / analyzing, media researchers use different approaches and datasets. In line with the project's proposal, we focused on historical analysis of television programmes, subtitles and newspapers. Accordingly, we divided the contextualization/analyzing phase into three parts: analyzing television in its historical context by means of comparison with newspapers (2.3.), analyzing the discourses of the television programmes by looking at the subtitles (2.4.) and watching the television programmes (2.5). We used the baseline technology, MeRDES, to have an initial indication for the findings, which helped us to imagine the next steps and the revision of the research questions. For features that are not present in MeRDES, we took examples from other interfaces. The scenario was written in three iterations, and concurrent feedback rounds with colleagues in computer sciences and media studies.

In the description of the scenario, we use the following typographical conventions: phrases that are underlined are functionalities of the interface. Some phrases are tagged with an asterisk (*); this indicates that these are not part of the original proposal, but we aim to implement them because the requirement analysis (D1) pointed out that these are desirable extensions.

¹ Bron, M., Van Gorp, J., Nack, F. F., de Rijke, M., Vishneuski, A. & de Leeuw, J.S. (2012). A Subjunctive Exploratory Search Interface to Support Media Studies Researchers. *SIGIR 2012: 35th international ACM SIGIR conference on research and development in information retrieval*. Portland, Oregon: ACM.

² Bron, M., Van Gorp, J., de Rijke, M. (under review) Media Studies Research in the Data-Drive Age. How Research Questions Evolve, *Journal of the American Society for Information Science and Technology*. Also see Van Gorp, J. (2013) Looking for what you are looking for: a media researcher's first search in a television archive. *VIEW: Journal of European Television History and Culture*, 2(1) online available at journal.euscreen.eu

2. Demonstration scenario

2.1. Setting the scene

Imagine that Corine wants to write an article on the historical representation of Russia on Dutch television. She wants to publish the article in the historical media studies journal *Historical Journal of Film, Radio and Television*. She uses QuaMeRDES for her research.

In general, she is empowered by QuaMeRDES to:

- formulate and investigate comparative research questions related to television and news papers / news events
- tailorize her results by making her own codes and by changing the settings of the visualisations and result lists
- create normalized statistics by taking into account the number of productions/articles per year, per topic and per missing field
- investigate the dialogues of television programmes and compare these with programme descriptions
- watch relevant television programmes*
- save*, export* and retrieve her results*

2.2. Exploring the television database

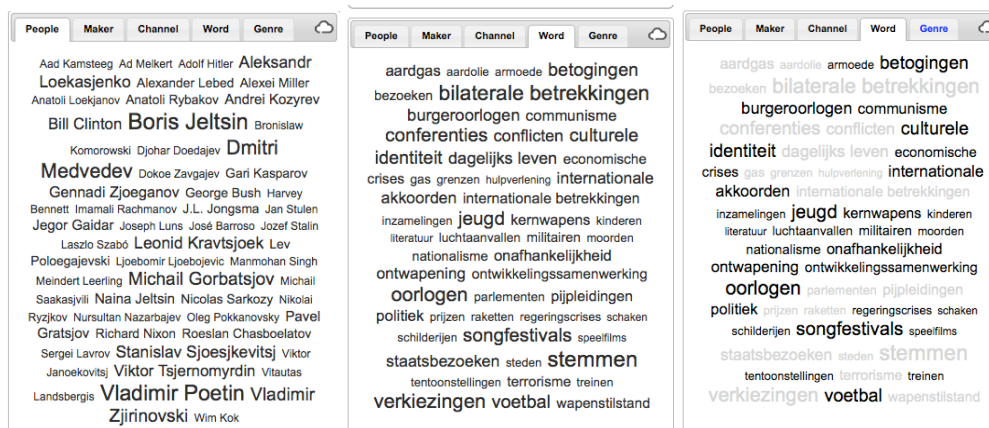
RQ1.How does the representation of Russia change over time in general terms?

RQ2.What are the general names, topics and genres associated with the representation of Russia on television?

Corine logs in with her e-mailaddress and password. First, she does an exploratory search on representation of Russia on television, in order to gain insight in the time frame on the one hand and key words and names related to television programmes about Russia on the other hand. On the timeline, she notices a peak in the early 1980s and a steady increase after 1991. She also discovers that names of politicians (Gorbachev, Putin) rank high in the name cloud. She also notices that words as protests, bilateral relations, and youth occur in the word cloud. She goes to the configuration panel and changes the settings of the word cloud from 20 words into 50 words, and then she notices also words such as soccer. She goes to the about page and reads that the word clouds' standard setting is 'parsimonious'. She goes to the configuration panel and unchecks the box. It turns out that for the word cloud 'parsimonious' is the

best setting.

Then, she creates a genre cloud. She sees that news rank highest, but also that music shows and sports are amongst the highest results. She wants to know whether the terms are genre-related, so she filters news out from the genre pie. By excluding news reports, other words pop up as most frequent: wars, cultural identity and soccer. It is cultural identity that triggers her attention.



Images 1-3. Name cloud (left), word cloud (middle) and filtered word cloud (right) (images taken from MerDES)

2.3. Analysing television in historical context

RQ3. To what extent and how is the representation of Russia on television related to news events?

Now, Corine wants to know whether the representation of Russia over time is related to the historical context. She can investigate this in two ways: (1) comparing fiction and non-fiction tv programmes and (2) comparing television with the historical context as represented in newspapers.

In a first round, she creates two codes for the television dataset in the query editor: non-fiction programmes and fiction programmes. By using boolean operators, she adds news, actualities, human interest, talk shows, sports, science to the code 'non-fiction programmes'. She adds films, drama, comedy, sitcom, animation to the code 'fiction programmes'. She compares both codes on a time line and in graphs.

In the second round, she does the same for fiction programmes and newspapers. She compares the code 'Russia in fictional tv-programmes' with Russia in newspapers, in order to see how both are related. She notices that the fiction programmes follow the peak in the newspapers.

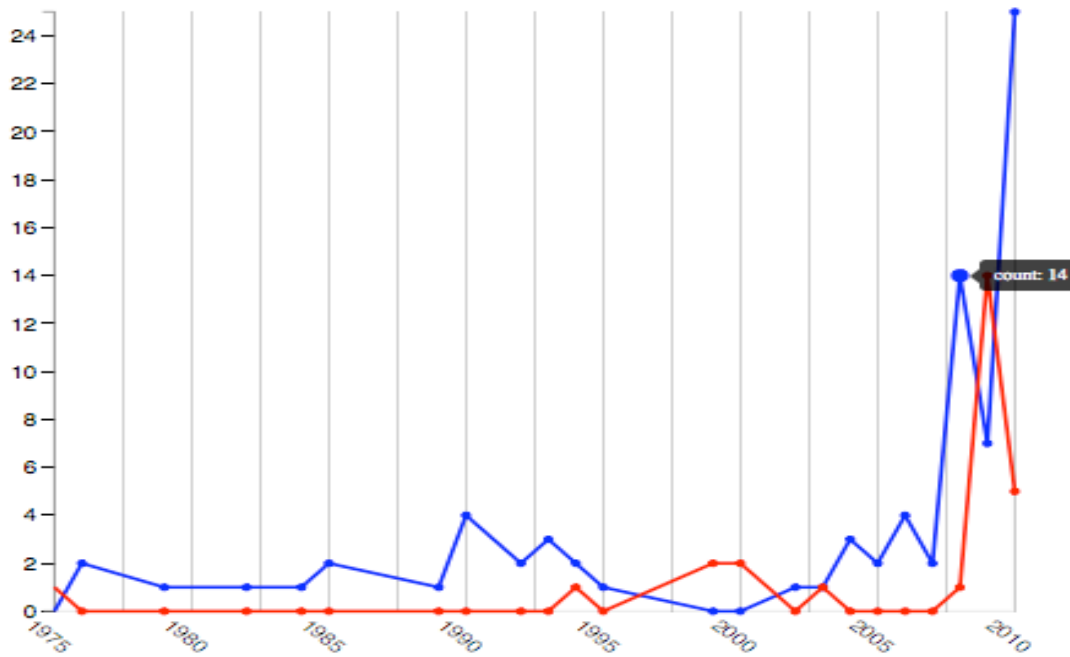


Image 4: Timeline comparison of query in two different datasets (image taken from MeRDES)

However, she realizes that she needs normalized statistics to take the increased number of programmes into account. She checks the box 'normalization by year', and observes that the sudden increase at the end of the timeline can be explained by an increase in production of tv-programmes rather than by an event. With the normalized statistics, she discovers that Gorbachov's glasnost and the collapse of the Soviet Union caused an increase in fiction programmes about Russia and Russians.

As last step, she wants to compare the different topics she discerned (i.e. crime, cultural identity) into a topic visualisation³ over time. She notices that there is an increase in topics on cultural identity in the 1980s and on crime in the 1990s. Therefore, she tracks the same topics in the newspapers. She notices that the topic-evolution in newspapers corresponds with this of the television programmes. However, the topic of cultural identity disappears in the television programmes of the 1990s, while in newspapers the topic has a steady increase.

³ In the project's proposal, we state that we will investigate possibilities to incorporate at least one additional visualisation. We are currently considering to implement both 'topic visualisation' and 'query in context' visualisation (see 2.4).

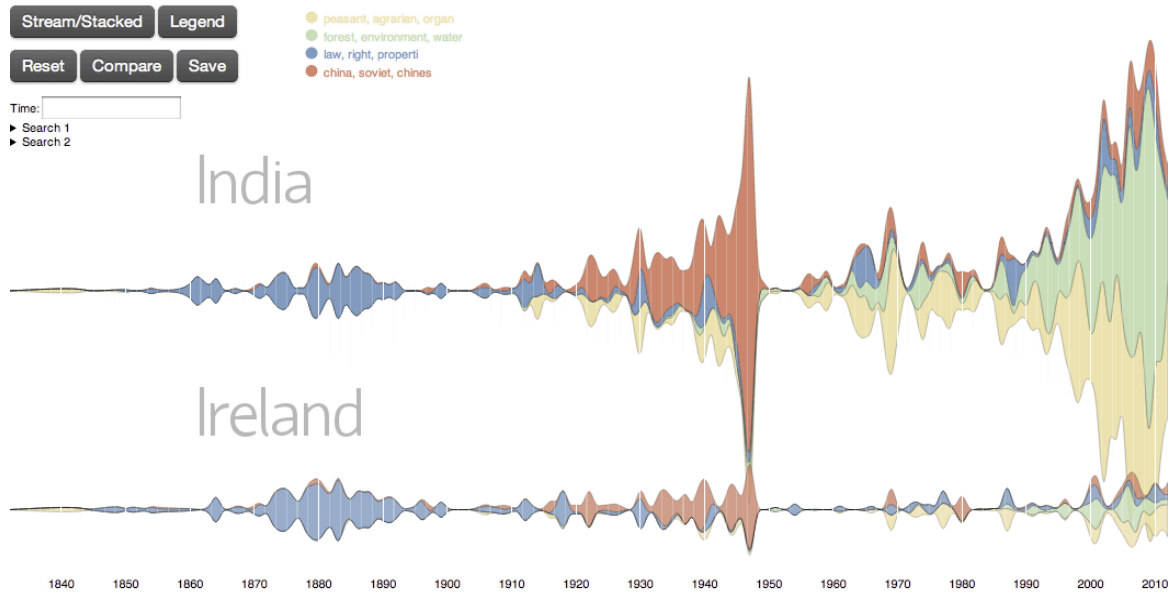


Image 5. Topic visualisation over time (image taken from Harvard's Paper Machines)

2.4. Comparing discourses

RQ4.In which context people in television programmes about Russia use the concept 'cultural identity'?

RQ5.Does it differ from the appearance of 'cultural identity' in programme descriptions?

RQ6.Does it differ from the appearance of 'cultural identity' in newspapers?

She remembers that in her first search 'cultural identity' popped up. 'It would be interesting', she thinks, 'to see how cultural identity is related to fiction and non-fiction'. Corine wants to investigate the subtitles of the programmes, in order to see in which context the people talk about 'cultural identity'. She creates clouds for the subtitles. She sees that crime has 213 hits and cultural identity 189 hits by hovering over the words. Then, she clicks on the word 'cultural identity' in the cloud and gets an overview of all sentences before and after the word cultural identity (Query in Context). By scanning all sentences, she finds that the words cultural identity in the television programmes are in 200 out of 213 cases mentioned together with 'Moscow', 'balalaika's' and 'vodka'. She bookmarks* the twenty programmes which rank high in the list.

She wants to take a meta-perspective and check whether there is a difference between the metadata description of the tv-programmes and the subtitles of the very same tv-programmes. She compares the code 'Russia and cultural identity' for the subtitle dataset with 'Russia and cultural identity' for the television dataset. And she repeats this action for 'Russia and cultural identity' for the subtitle dataset and the newspaper dataset. She checks the box 'missing data'⁴ and sees that out of 1.700 programmes

⁴ One of the options to normalize the statistics is to take the number of programmes into account which have empty programme description fields (= 'missing data').

produced, only 254 have subtitles. 1.540 tv-programmes do have extensive metadata-descriptions. She saves her results*.

#	Examples	Morphosyntax	Resume/Def	Domain	SUMOntology	Synonyms	Relations	Tagged
1	op de televisie	n-de-t	ontvangststelsel voor tv...	natur bouwk...	Device	buis kastje kijkbuis	toestel apparaat	0
2	van de televisie	n-de-t	systeem van televisieom...	media nat ...	Device Tele...	omroepbestel	0	
3	Dat heb ik op televisie gehoord.	n-t	programma dat op de tv ...	media	Disseminat...	televisieprogramma televisie-uit...	programma uitzending	0

#	tfeL	Left	Sense	Word	Right
1	nee...	ien hoor ik dus nooit iets . hum . geen uhm ... ik weet niet of ze ooit een		televisie	aan hebben . of dat ze ooit muziek aan hebben . en zij hebben van die uh
2	djit...	in een compleet donkere ruimte de schittering van de TV . oh hij zit altijd		televisie	te kijken ? ja . maar dat hoor je niet in jouw slaapkamer ? dat hoor ik
3	ed ...	jna dertig of zo ? ik weet niet hoe oud ie is . was pas een quizvraag op de		televisie	uh ja 'a één of andere vreselijke lullige quiz maar ... van uh wat is het
4	ed j ...	n . uh ve'a vijfendertig gulden ja . voor twee meter . maar daarvoor bij de		televisie	daar komen wel vier of vijf schilderijtjes . en bovendien'x xxx op andere h
5	a j te doen hebben . en dat vindt ze heerlijk . ja . ja		televisie	aan en breien . want gewoon televisiekijken doet ze niet . ze moet'x altijd
6	po r ...	t over heb . nu moet 'k in één keer denken aan die kuikentjes uh gister op'		televisie	die bij uh k'a met m'a met z'n vieren vijen tegelijkertijd in één hand kwa
7	po l ...	ien joh . gewoon'x . waar waar is dat ook alweer ? ja je zal 't vast wel op		televisie	hebben gezien . want 't is ... nee we hebben dat gezien in uhm in Lissabon
8	po ...	s gewoon heel veel uh ... uhm ... hier veel veel ... gewoon weer wat ook op		televisie	is achter elkaar gezet als het ware . oh ja . ja is ideaal en dan uh heb ik
9	ed ...	ndacht krijgt . vanmorgen'z was ze toch niet goed te pas . nee . mwa als de		televisie	maar aan was dan was 't dan nog wel aardig uh ... mm-hu . maar ze heeft liet
10	ed ...	ijk van uh hoe dat hoe dat loopt . is dat dan uh zo'n soort uh wat je op de		televisie	hebt ? Buiten'a Buitenhof of zo ? ja . ja en daar tussendoor ja nou gewoon
11	ed r ...	doet geheimzinniger dan in de Tweede Kamer . want die vertellen 't voor de		televisie	. ggg . ja ik zeg maar ook na twee drie uur dus laat me
12	po ...	dat was uh dat was uh vanmiddag nog op uh pff ik weet niet was uh was op .		televisie	ik had a'a toen lag ik eventjes op de bank ik denk met vijf uh 't vijf-uur-
13	po ...	natuur en zo is wel heel mooi natuurlijk want dat ja dat heb je wel 'ns op		televisie	gezien en zo . ja 't heet de Bloemenriviera . ik dacht altijd dat dat Zuid-
14	po ...	is best'x spannend 't is natuurlijk wel zeg en je zit'a je ziet van alles op		televisie	en je leest heel erg van alles in de krant . spannend natuurlijk . ja . je
15	ed r ...	ag niet . 't moet serieus zijn . ja . was een uhm meneer en die zat voor de		televisie	en toen zag ie in zag ie uh de uhm hoorde ie eigenlijk . hij moet serieus z
16	po ...	hij zag 't juist niet en daarom hoorde ie 't . dat De Bus wat later uh op .		televisie	was als dat 't aangekondigd was in de televisiegids . en toen drukte ie de
17	ed ...	was als dat 't aangekondigd was in de televisiegids . en toen drukte ie de		televisie	uit . en zeggen ze van'x waarom deed je dat nou ? ja zeg ie 't
18	ezn ...	aar hij moe'a hij moet er even voor'x gezorgd ook wa'u we 't geld voor onze		televisie	betalen aan hem . ja . moet ik 't voorschieten ? ja graag . ik wil 't dus
19	a's ...	vorige week was de eerste aflevering . was heel stom . ja . s'a .		televisie	of zo heet 't of niet ? ja maar 't was echt ... het slaat helemaal nergens
20	tad ...	or m'n politico-logische'u achtergronden . zekers zekers zeker . maar dat		televisie	dat is echt niet de moeite waard . want ik was echt gewoon serieus opgebevl
21	...jwie heeft dan die prijs verhoogd ? hij ? mmm ja nee hij		televisie	? hij heeft ervoor gezorgd in ieder geval dat het omhoog ging . dan heb
22	nee ...	ja . en bij uh be'a uh be'a ook bij één trouwens niet . heb g'a ook geen		televisie	. oh nee . en bij uh één is 't ook dat je dat ze'x
23	nee ...	een is 't ook dat je dat ze'x heel lastig met het karretje . geen		televisie	? 'a dat past allemaal net . en dan moeten ze eigenlijk eerst zelf d'rin
24	hu t ...	ken dan tikt ie ook twee drie flesjes bier weg . dan 's avonds uh bij 't uh		televisie	kijken een flesje wijn d'rbij . dat tikt ie ook zo uh makkelijk weg . zo in
25	po ...	met uh met Victor Schipper dat weet ik nog wel . die was laatst trouwens op		televisie*x	. oh ja ? en uh mocht ik zelf een onderwerp kiezen . en dat
26	ed ...	de wieg al ggg ggg . ja . daar kun je mee opschieten . de te'a de		televisie	ggg . twee nummers uit hun hoofd leren . s'a zulk'x uh zulk soort figuren w
27	ed r ...	h ook zeer uh passend moet ik zeggen . ja . ja ja d'r werd ook door de		televisie	opnamen gemaakt . ik wat ik ook heel mooi vond dat waren die uh ... was hel
28	ed ...	in de gids kijken ? waarom ? xxx ggg . nou misschien is d'r wel iets op de		televisie	. ja uh X-files*v . X-files*v bestaat dat nog ? ja . zijn 't herhalingen
29	po ...	hele tijd zo van ... die'v . en waarom wil je je tieten niet laten zien op		televisie	? ja dat wil ik gewoon niet . ja waarom wil je dat dan niet

Image 6. Example of query in context. We can use it to contextualize the words of the subtitle word-cloud (image taken from Dutch semcor)

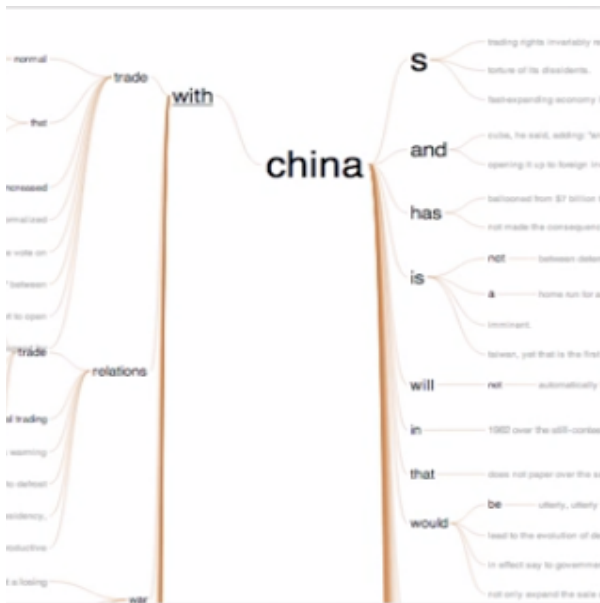


Image 7. Example of query in context (image taken from Berkeley's Wordseer 3.0.)

2.5. Watching clips

RQ7.How is cultural identity visually represented in the television programmes?

RQ8.To what extent and how do the findings of the historical analysis and the discourse analysis correspond with those of the actual programmes?

Once she has gained insight in some trends, she wants to watch some programmes* to see whether and how the results of her previous searches actually match with the television programmes. She selects one programme for every decennium on the basis of her bookmark list*.

2.6. Writing up the research and retrieving previous visualisations

One month later, Corine starts to write up her research article. She visits her personal page* in the interface where she saved all her previous searches and visualisations, including the dates that these were done, and downloads* them to her computer.