Exploiting Program Guides for Contextualisation

Lotte Belice Baltussen
Netherlands Institute for Sound and Vision
Hilversum, The Netherlands
Email: lbbaltussen@beeldengeluid.nl

Themistoklis Karavellas
Netherlands Institute for Sound and Vision
Hilversum, The Netherlands
Email: tkaravellas@beeldengeluid.nl

Roeland J.F. Ordelman
University of Twente,
Enschede, The Netherlands
Email: roeland.ordelman@utwente.nl

***Abstract*—Archives of cultural heritage organisations typically consist of collections in various formats (e.g. photos, video, texts) that are inherently related. Often, such disconnected collections represent value in itself but effectuating links between ’core’ and ’context’ collection items in various levels of granularity could result in a ’one-plus-one-makes-three’ scenario both from a contextualisation perspective (public presentations, research) and access perspective. A key issue is the identification of contextual objects that can be associated with objects in the core collections, or the other way around. Traditionally, such associations have been created manually. For most organizations however, this approach does not scale. In this paper, we describe a case in which a semi-automatic approach was employed to create contextual links between television broadcast schedules in program guides (context collection) and the programs in the archive (core collection) of a large audiovisual heritage organisation.**

***Index Terms*—Contextualisation, Cultural Heritage, Data Analysis, Linked Collections**

I. Introduction

The collections of heritage organisations such as libraries, museums and archives typically consist of collections in various formats that are inherently related. For instance, a National Library possesses books as a physical object and digital representation, but possibly also private collections of the authors of these books such as letters, notes or even objects. An art museum may have paintings, but also the original pamphlets used for previous exhibitions featuring these paintings [1]. An audiovisual archive might have a collection of broadcast video, and also television program guides or photos taken during the production of a program on the set. Although each collection represent value in itself, making explicit links between these ’core’ and ’context’ collection items in various levels of granularity increases their value both from a contextualization perspective (public presentations, research) and access perspective. Access scenarios are for example the use of information from contextual sources to enrich content descriptions of core data to improve ’searchability’, or providing new entry points to either core or context collections via the establishment of explicit (hyper)links between items from the collections. For instance, the program descriptions in program guides can be used as *metadata* for audiovisual broadcast content, or end-users could be pointed towards audiovisual sources while browsing the collection of program guides.

A key issue, either from a contextualization perspective or an access perspective, is the identification of contextual media objects that can be associated with media objects in the core collections, or the other way around. Traditionally, such associations have been created manually by media archivists or collection specialists. For most organizations however, this approach does not scale. For large-scale contextual interlinking a (semi-) automatic approach is required: algorithms detect relations between items and suggest links that are validated manually by humans (e.g. archivists).

We consider a use case of an audiovisual archive with a large collection of broadcast video data, and a digitized collection of program guides representing four decades of television history. The guides contain TV and radio broadcast schedules, articles, pictures, advertisements and other elements that provide insight in the cultural-historical context of the broadcast videos. The use case aims at the automatic identification of program guide material that can be linked to the broadcasts in the archive, and by doing so, to contextualise the core collection with context data. Practically, the program guide material will be used as metadata to improve ’searchability’ of archive video. Moreover, via an online program guide browser, direct access from the guides to related archival content will be provided.

In this paper, we describe how our work relates to other developments in the Digital Humanities domain, our approach for extracting and matching data from OCR files to the Netherlands Institute for Sound and Vision’s (NISV) av-archive and our lessons-learned.

II. Related Work

In the LODLAM community[1](Linked Open Data in Libraries, Archives and Museums) [2], [3] many activities are undertaken to connect collections from heritage institutions to both internal and external information. Linked Open Data technologies are used to support the process of creating connections. For example, the *Free Your Metadata*[2] initiative provides easy-to-use tools for making metadata part of the Linked Data cloud and organizes hands-on workshops.

As part of the CATCH Programme, together with the Intelligent Systems Lab Amsterdam at the University of Amsterdam, the Centre for Television in Transition at Utrecht University, we developed the tool CoMeRDa (Contextualizing Media Research Data)[3]). It is an aggregated system that allows searching full-text across diverse data collections including television catalogue metadata, newspaper articles, photo collections, program guides and a specialized wiki. The interface combines information from the archives’ core collection (catalogue info) with its context collections. In CoMeRDa, it is possible to use a record as a search query across the

[1] http://lodlam.net/
[2] http://freeyourmetadata.org/
[3] http://vps46235.public.cloudvps.com/bridge/tools/the-comerda-tool/

978-1-5090-0048-7/15/$31.00 ©2015 IEEE

collections, but no explicit Linked Open Data or other linking technologies were used to make relations between records and collections explicit. [4]

Within the scope of the BBC Genome project [4] 350,000 pages of Radio Times program guides from 1922-2009 were digitised. Optical Character Recognition (OCR) software was used to automatically convert the characters on the digitised pages into machine-encoded text. This OCR process not only makes it possible to provide full-text search, but also to segment the information on the pages into distinct chunks. By analysing the mark-up and layout of the program schedules, the information about programs aired on television and radio were automatically turned into a massive database, listing a total of almost 4.5 million BBC program records. These records contain information on the program title, description, date, channel, cast, airing time, related to the specific page of the guide the information came from.

The project described in this paper took the approaches described above one step further by linking the extracted data of the context collection of the guides to the core av-collection of the archive, and explores how to optimise the quality of the extracted output and detected links.

III. APPROACH

A. *Data collection*

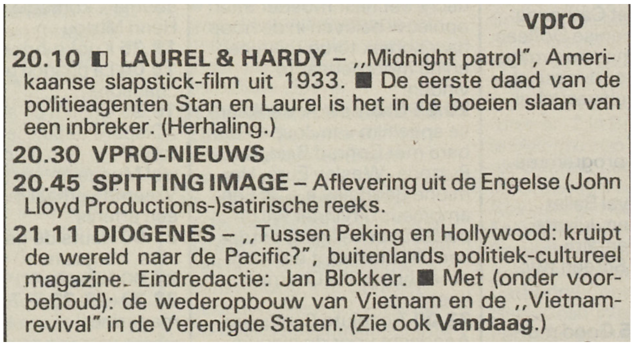
Our corpus consists of 330 guides of six public broadcasters[5], spanning four decades: 1951-1986. We didn't scan all guides from this period, but used intervals of five years, and for each year scanned two months worth of issues. For example, for 1951 we scanned the guides of all six broadcasters from January and September, for 1956 we did the same for the February and October, and so on. In total, 25,000 scanned pages were part of our corpus.

For each page, the OCR software Abbyy FineReader 10 was used to convert the image scans into a computer readable format (XML). We used a parsing script[6] developed specifically for the project to segment the XMLs of the broadcast schedules of TV programs of public broadcasters, in order to create a database of broadcasts that spans four decades of Dutch television history. For each record in the database (representing one program) we tried to capture the following structured information: date, start time of the broadcast, end time of the broadcast, channel, broadcaster name, title and description. In a next step, we mapped the extracted titles with those in the AV catalogue of NISV, in order to make the match between specific programs in the archive and the program guide information about them. This results in linking the previously analogue, paper contextual collections and the core av-archive.

B. *Parsing the OCR*

There is more information in the guides than TV schedules, such as advertisements, articles and photos. Although these additional data provide great contextual information as well, we focused only on extracting descriptions regarding Dutch public TV programs. This process starts with identifying the correct information to parse by analysing the OCR output captured in a raw OCR file. In this raw file, different parameters are used to described each detected character: for example the coordinate of the character on the page, confidence score, font type, font size, the position of the character within a word, and if it is bold or italic. Our data extraction algorithm groups separate characters into words. A group of words form a line and a group of lines form a block. The block of text is analysed to determine whether it contains TV program schedule information.

A block of text is checked for elements that typically describe TV programs, such as temporal elements representing start and end times, a date, and broadcaster and TV channel names from a predefined list. If temporal elements and a title and description are found, we assume that the block contains information about one single program. The first temporal element in this block is considered to be the start time. The rest of the text is parsed to detect the other elements: end time, broadcaster name, and the title and program description. The channel name and date are usually found outside of the text block, at the top of a page and are extracted separately. (see Fig. 1). Following this approach, each text block describing an individual TV program populates an XML record. Each attribute gets its own place within this record, including the page of the guide on which the text block can be found. After all pages in a guide are parsed and all detected TV program text blocks are examined, the extracted records are saved in an XML file representing the entire guide.



vpro

20.10 ◧ LAUREL & HARDY – ,,Midnight patrol", Amerikaanse slapstick-film uit 1933. ■ De eerste daad van de politieagenten Stan en Laurel is het in de boeien slaan van een inbreker. (Herhaling.)

20.30 VPRO-NIEUWS

20.45 SPITTING IMAGE – Aflevering uit de Engelse (John Lloyd Productions-)satirische reeks.

21.11 DIOGENES – ,,Tussen Peking en Hollywood: kruipt de wereld naar de Pacific?", buitenlands politiek-cultureel magazine. Eindredactie: Jan Blokker. ■ Met (onder voorbehoud): de wederopbouw van Vietnam en de ,,Vietnamrevival" in de Verenigde Staten. (Zie ook **Vandaag**.)

Fig. 1. Example of a TV broadcast schedule of four progammes, containing the start times, titles and descriptions. The names of the cast also mentioned. In this case, the date and channel name are found elsewhere on the page

In total, 990,000 distinct broadcasts were identified within the corpus of 330 guides containing 25,000 pages. It needs to be noted that the bulk of these broadcasts overlap, as guides –originally published by different broadcasters– represent the same years and weeks. Also, this set contains programs that were broadcast more than once (reruns). A sample set was manually checked for accuracy of the extracted elements, which showed that 91.6 percent of the title and description elements are (almost) complete and correct and that in almost all cases the start time of a program was extracted accurately. The elements channel, broadcaster and date were however only parsed correctly in around 50-55 percent of the cases.

[4] http://genome.ch.bbc.co.uk/

[5] In the Netherlands, the public broadcast system consists of many different broadcasters, each targeting a societal group (e.g. Protestants, Catholics). We scanned guides from the following broadcasters: AVRO,KRO,NCRV,TROS,VARA,VPRO

[6] https://github.com/beeldengeluid/tvguide-segmenter

C. Main parsing issues

We encountered a number issues which made it difficult to parse information correctly. The main issues are:

- Incorrect OCR output: OCR-output typically has flaws. Due to ink bleeding, rips and tears in the paper, ink fading or poor print quality some characters in the OCR output were incorrectly detected.
- Text flow order: the guides present TV schedules in different ways. Some guides present them in neatly defined vertical columns per broadcasters, whereas others use a horizontal presentation. This makes it hard to identify to which channel a text block belongs and to identify when the text in a text block ends on the bottom of a column and continues at the top of the next column (see left page in Fig. 2).
- Layout identification: each guide has its own graphic design and way to arrange elements on a page. Each guide has its own way of presenting information: some print the start and end times of programmes in Italic, some in bold, others always print program titles in bold and font size 16, in some guides the date is presented at the top left of a page and so on. Usually, this styling of elements differs per broadcaster, per year.
- Page selection for segmentation: we only wanted to parse TV program schedules from Dutch public broadcasters. To this end, we had to create specific lists of TV broadcaster and channel names in order to identify the correct pages to parse.
- Date parsing: in general, the day of the program schedule is mentioned at the top of the page. However, many formats are used to describe the date e.g. ”Friday 13 May” or ”12 Jan”. This makes it very hard to convert the date –if detected by the parsing script at all– to a unified format.

The most important action undertaken in the project to improve parsing accuracy was creating so-called configuration files. These files convey information about the text flow order and layout of information of a certain guide, e.g. the location of the date, the font size of titles and which elements were presented in bold. Configuration files were manually created for each guide per year and provided to the parsing script. Although it took a couple of days to create these configuration files, the accuracy output of the script improved dramatically, in some cases resulting in the detection of three times as many individual program blocks on a single page.

D. Matching parsed results with television programs

After the TV program schedules were parsed, links were identified between the extracted program guide records and the corresponding programs in NISV's catalogue. This was done by creating a list of program titles in the catalogue corresponding to the date range of each parsed program guide. So, if a guide represents information from 8 January 1961 to 14 January 1961, only programs in the catalogue from this date range are considered in the matching process. When a title was detected in the description of a parsed record, this was recorded as a match. This process is however not 100 percent accurate:

- The date extracted from the guide was not always correct or not detected at all.
- The matching script also allowed partial matching. This means that a title like ”Jeugdjournaal” (Children's News) also matches the title ”Journaal” (News).
- Reruns of programs appear in the program schedule of a guide, but in the NISV archive generally only the first, original broadcasts are stored. This can result in a false positive match when a program is rerun within the date range of a single guide.

The total *match potential* was $6,462$. In other words, for the dates of the parsed guides, there are $6,462$ individual television programs in NISV's archive. Of these programmes, $2,611$ were matched. Because we scanned six different guides from the same weeks, in many cases more than one match was found per program, resulting in a total of $6,224$ matches (see Table I). This amount may seem low, considering that $990,000$ records were found. However, there are large gaps in the NISV archive for the period of the parsed guides (330 guides from 1951-1986). The reason is that it only became possible to archive all broadcasts in the 2000s, when fully digital workflows were implemented. Before, tapes were sent back and forth between broadcasters and the archive, and a selection policy was employed as it was too costly to archive everything. Moreover, sometimes other programs besides Dutch public television programs were parsed by our script, such as radio broadcasts and programs from foreign channels.

Matches per program	Occurrences	Total matches
1	1,147	1,147
2	432	864
3	288	864
4	422	1,688
5	275	1,375
6	47	282
TOTAL	**2,611**	**6,220**

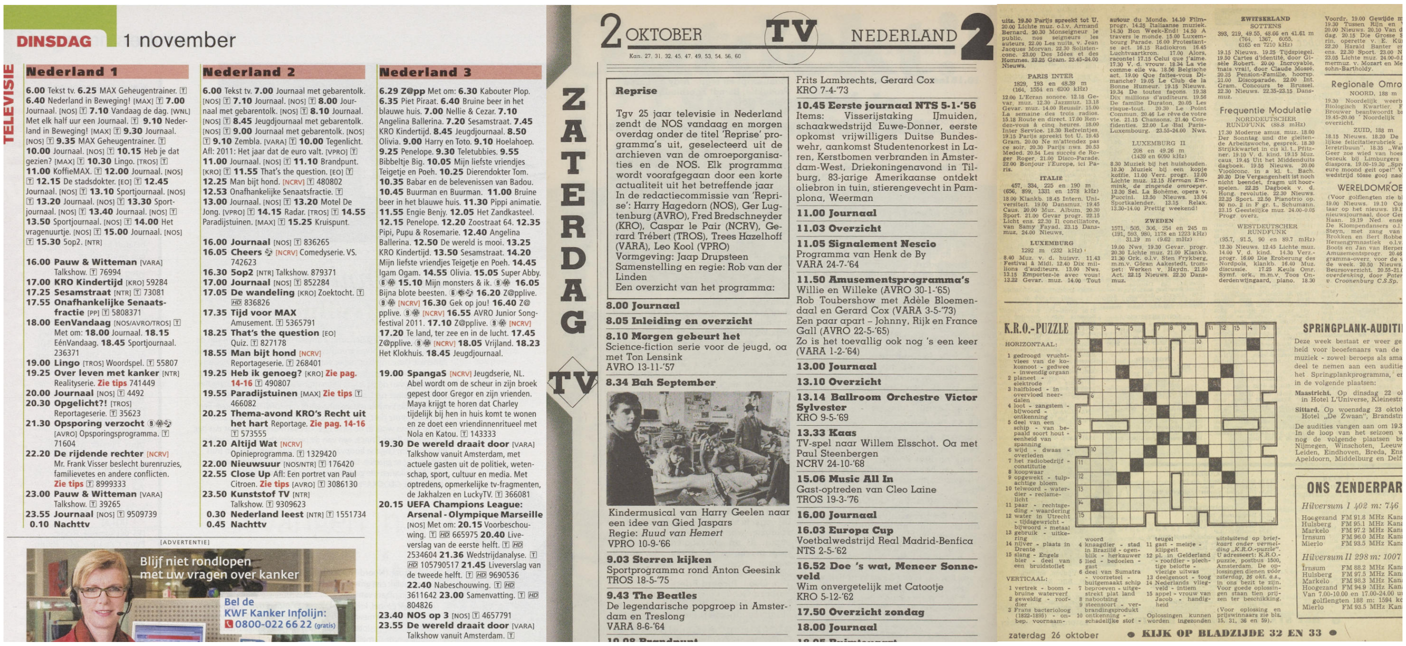
TABLE I. NUMBER OF MATCHES PER PROGRAM, RANGING FROM A MATCH FROM 1 GUIDE TO 6 GUIDES

A representative sample of the matched records was checked manually for accuracy. We found that 89.2 percent of the matches are accurate, and that for 6.6 percent of the inaccurate matches the right program was matched, but not the right episode. The latter was due to the date being incorrect or not present in the parsed output. This means that in 4.2 percent of the cases the match was completely inaccurate. However, the match to the guide itself is always correct, since the broadcast date of the program always falls within the date range of the guide.

E. Extracting thesaurus terms

For each parsed title and description element within an identified text block, Named Entity Recognition was used to extract terms from the archive's thesaurus[7], which contains terms relevant to the NISV archive. For 63.8 percent of the $6,224$ total matches, one or more terms were extracted. 73

[7] http://openskos.beeldengeluid.nl/api



DINSDAG 1 november

TELEVISIE

Nederland 1

6.00 Tekst tv. **6.25** MAX Geheugentrainer. 🅃 **6.40** Nederland in Beweging! [MAX] 🅃 **7.00** Journaal. [NOS] 🅃 **7.10** Vandaag de dag. [WNL] Met elk half uur een Journaal. 🅃 **9.10** Nederland in Beweging! [MAX] 🅃 **9.30** Journaal. [NOS] 🅃 **9.35** MAX Geheugentrainer. 🅃 **10.00** Journaal. [NOS] 🅃 **10.15** Heb je dat gezien? [MAX] 🅃 **10.30** Lingo. [TROS] 🅃 **11.00** KoffieMAX. 🅃 **12.00** Journaal. [NOS] 🅃 **12.15** De stadsdokter. [EO] 🅃 **12.45** Journaal. [NOS] 🅃 **13.10** Sportjournaal. [NOS] 🅃 **13.20** Journaal. [NOS] 🅃 **13.30** Sportjournaal. [NOS] 🅃 **13.40** Journaal. [NOS] 🅃 **13.50** Sportjournaal. [NOS] 🅃 **14.00** Het vragenuurtje. [NOS] 🅃 **15.00** Journaal. [NOS] 🅃 **15.30** 5op2. [NTR]

16.00 Pauw & Witteman [VARA] Talkshow. 🅃 76994
17.00 KRO Kindertijd [KRO] 59284
17.25 Sesamstraat [NTR] 🅃 73081
17.55 Onafhankelijke Senaatsfractie [PP] 🅃 5808371
18.00 EenVandaag [NOS/AVRO/TROS] 🅃 Met om: **18.00** Journaal. **18.15** EénVandaag. **18.45** Sportjournaal. 236371
19.00 Lingo [TROS] Woordspel. 🅃 55807
19.25 Over leven met kanker [NTR] Realityserie. Zie tips 741449
20.00 Journaal [NOS] 🅃 4492
20.30 Opgelicht?! [TROS] Reportageserie. 🅃 35623
21.30 Opsporing verzocht ⑨ [AVRO] Opsporingsprogramma. 🅃 71604
22.20 De rijdende rechter [NCRV] Mr. Frank Visser beslecht burenruzies, familievetes en andere conflicten. Zie tips 🅃 8999333
23.00 Pauw & Witteman [VARA] Talkshow. 🅃 39265
23.55 Journaal [NOS] 🅃 9509739
0.10 Nachttv

Nederland 2

6.00 Tekst tv. **7.00** Journaal met gebarentolk. [NOS] 🅃 **7.10** Journaal. [NOS] 🅃 **8.00** Journaal met gebarentolk. [NOS] 🅃 **8.10** Journaal. [NOS] 🅃 **8.45** Jeugdjournaal met gebarentolk. [NOS] 🅃 **9.00** Journaal met gebarentolk. [NOS] 🅃 **9.10** Zembla. [VARA] 🅃 **10.00** Tegenlicht. Afl: 2011: Het jaar dat de euro valt. [VPRO] 🅃 **11.00** Journaal. [NOS] 🅃 **11.10** Brandpunt. [KRO] 🅃 **11.55** That's the question. [EO] 🅃 **12.25** Man bijt hond. [NCRV] 🅃 480802 **12.53** Onafhankelijke Senaatsfractie. 🅃 **13.00** Journaal. [NOS] 🅃 **13.20** Motel De Jong. [VPRO] 🅃 **14.15** Radar. [TROS] 🅃 **14.55** Paradijstuinen. [MAX] 🅃 **15.25** Kruispunt.

16.00 Journaal [NOS] 🅃 836265
16.05 Cheers [NCRV] Comedyserie. VS. 742623
16.30 5op2 [NTR] Talkshow. 879371
17.00 Journaal [NOS] 🅃 852284
17.05 De wandeling [KRO] Zoektocht. 🅃 HD 836826
17.35 Tijd voor MAX Amusement. 🅃 5365791
18.25 That's the question [EO] Quiz. 🅃 827178
18.55 Man bijt hond [NCRV] Reportageserie. 🅃 268401
19.25 Heb ik genoeg? [KRO] Zie pag. 14-16 🅃 490807
19.55 Paradijstuinen [MAX] Zie tips 🅃 466082
20.25 Thema-avond KRO's Recht uit het hart Reportage. Zie pag. 14-16 🅃 573555
21.20 Altijd Wat [NCRV] Opinieprogramma. 🅃 1329420
22.00 Nieuwsuur [NOS/NTR] 🅃 176420
22.55 Close Up Afl: Een portret van Paul Citroen. Zie tips [AVRO] 🅃 3086130
23.50 Kunststof TV [NTR] Talkshow. 🅃 9309623
0.30 Nederland leest [NTR] 🅃 1551734
0.45 Nachttv

Nederland 3

6.29 Z@pp Met om: **6.30** Kabouter Plop. **6.35** Piet Piraat. **6.40** Bruine beer in het blauwe huis. **7.00** Nellie & Cezar. **7.10** Angelina Ballerina. **7.20** Sesamstraat. **7.45** KRO Kindertijd. **8.45** Jeugdjournaal. **8.50** Olivia. **9.00** Harry en Toto. **9.10** Hoelahoep. **9.25** Penelope. **9.30** Teletubbies. **9.55** Bibbeltje Big. **10.05** Mijn liefste vriendjes Teigetje en Poeh. **10.25** Dierendokter Tom. **10.35** Babar en de belevenissen van Badou. **10.45** Buurman en Buurman. **11.00** Bruine beer in het blauwe huis. **11.30** Pippi animatie. **11.55** Engie Benjy. **12.05** Het Zandkasteel. **12.15** Penelope. **12.20** Zoostraat 64. **12.35** Pipi, Pupu & Rosemarie. **12.40** Angelina Ballerina. **12.50** De wereld is mooi. **13.25** KRO Kindertijd. **13.50** Sesamstraat. **14.20** Mijn liefste vriendjes Teigetje en Poeh. **14.45** Igam Ogam. **14.55** Olivia. **15.05** Super Abby. ⑥ **15.10** Mijn monsters & ik. ⑨ **16.05** Bijna blote beesten. ⑥ **16.20** Z@pplive. ⑨ [NCRV] **16.30** Gek op jou! **16.40** Z@pplive. ⑨ [NCRV] **16.55** AVRO Junior Songfestival 2011. **17.10** Z@pplive. ⑨ [NCRV] **17.20** Te land, ter zee en in de lucht. **17.45** Z@pplive. ⑨ [NCRV] **18.05** Vrijland. **18.23** Het Klokhuis. **18.45** Jeugdjournaal.

19.00 SpangaS [NCRV] Jeugdserie, NL. Abel wordt om de scheur in zijn broek gepest door Gregor en zijn vrienden. Maya krijgt te horen dat Charley tijdelijk bij hen in huis komt te wonen en ze doet een vriendinnenritueel met Nola en Katou. 🅃 143333
19.30 De wereld draait door [VARA] Talkshow vanuit Amsterdam, met actuele gasten uit de politiek, wetenschap, sport, cultuur en media. Met optredens, opmerkelijke tv-fragmenten, de Jakhalzen en LuckyTV. 🅃 366081
20.15 UEFA Champions League: Arsenal - Olympique Marseille [NOS] Met om: **20.15** Voorbeschouwing. 🅃 HD 665975 **20.40** Liveverslag van de eerste helft. 🅃 HD 2534604 **21.36** Wedstrijdanalyse. 🅃 HD 105790517 **21.45** Liveverslag van de tweede helft. 🅃 HD 9690536 **22.40** Nabeschouwing. 🅃 HD 3611642 **23.00** Samenvatting. 🅃 HD 804826
23.40 NOS op 3 [NOS] 🅃 4657791
23.55 De wereld draait door [VARA] Talkshow vanuit Amsterdam. 🅃

[ADVERTENTIE]

Blijf niet rondlopen met uw vragen over kanker

Bel de KWF Kanker Infolijn: 0800-022 66 22 (gratis)

ZATERDAG TV

2 OKTOBER TV NEDERLAND 2

Kan. 27, 31, 32, 45, 47, 49, 53, 54, 56, 60

Reprise

Tgv 25 jaar televisie in Nederland zendt de NOS vandaag en morgen overdag onder de titel 'Reprise' programma's uit, geselecteerd uit de archieven van de omroeporganisaties en de NOS. Elk programma wordt voorafgegaan door een korte actualiteit uit het betreffende jaar
In de redactiecommissie van 'Reprise': Harry Hagedorn (NOS), Ger Lugtenburg (AVRO), Fred Bredschneyder (KRO), Caspar le Pair (NCRV), Gerard Trébert (TROS), Trees Hazelhoff (VARA), Leo Kool (VPRO)
Vormgeving: Jaap Drupsteen
Samenstelling en regie: Rob van der Linden
Een overzicht van het programma:

8.00 Journaal

8.05 Inleiding en overzicht

8.10 Morgen gebeurt het
Science-fiction serie voor de jeugd, oa met Ton Lensink
AVRO 13-11-'57

8.34 Bah September

Kindermusical van Harry Geelen naar een idee van Gied Jaspars
Regie: *Ruud van Hemert*
VPRO 10-9-'66

9.03 Sterren kijken
Sportprogramma rond Anton Geesink
TROS 18-5-'75

9.43 The Beatles
De legendarische popgroep in Amsterdam en Treslong
VARA 8-6-'64

10.08 Brandpunt

Frits Lambrechts, Gerard Cox
KRO 7-4-'73

10.45 Eerste journaal NTS 5-1-'56
Items: Visserijstaking IJmuiden, schaakwedstrijd Euwe-Donner, eerste opkomst vrijwilligers Duitse Bundeswehr, aankomst Studentenorkest in Laren, Kerstbomen verbranden in Amsterdam-West, Driekoningenavond in Tilburg, 83-jarige Amerikaanse ontdekt oliebron in tuin, stierengevecht in Pamplona, Weerman

11.00 Journaal

11.03 Overzicht

11.05 Signalement Nescio
Programma van Henk de By
VARA 24-7-'64

11.50 Amusementsprogramma's
Willie en Willeke (AVRO 30-1-'65)
Rob Toubershow met Adèle Bloemendaal en Gerard Cox (VARA 3-5-'73)
Een paar apart – Johnny, Rijk en France Gall (AVRO 22-5-'65)
Zo is het toevallig ook nog 's een keer (VARA 1-2-'64)

13.00 Journaal

13.10 Overzicht

13.14 Ballroom Orchestre Victor Sylvester
KRO 9-5-'69

13.33 Kaas
TV-spel naar Willem Elsschot. Oa met Paul Steenbergen
NCRV 24-10-'68

15.06 Music All In
Gast-optreden van Cleo Laine
TROS 19-3-'76

16.00 Journaal

16.03 Europa Cup
Voetbalwedstrijd Real Madrid-Benfica
NTS 2-5-'62

16.52 Doe 's wat, Meneer Sonneveld
Wim onvergetelijk met Catootje
KRO 5-12-'62

17.50 Overzicht zondag

18.00 Journaal

18.05 Ruimtevaart

uitz. 19.50 Parijs spreekt tot U. 20.00 Lichte muz. o.l.v. Armand Bernard. 20.30 Monseigneur le public, nos seigneurs les auteurs. 22.00 Les nuits, v. Jean Jacques Morvan. 22.30 Solistenconc. 23.00 Des Idées et des Hommes. 23.25 Gram. 23.45-24.00 Nieuws.

PARIS INTER
1829, 193 en 48.39 m
(164, 1554 en 6200 kHz)

12.00 L'Ecran sonore. 12.18 Gevar. muz. 12.30 Jazzmuz. 13.18 Gevar. muz. 14.00 Reussir. 15.00 La semaine des trois radios. 15.18 Route en direct. 17.00 Rendez-vous à cinq heures. 18.00 Inter Service. 18.30 Refreintjes. 19.15 Parijs spreekt tot U. 19.45 Gram. 20.00 Ne m'attendez pas ce soir. 20.30 Parijs nws. 20.53 Meded. 20.54 Les succès de Roger Roger. 21.00 Disco-Parade. 22.00 Bonjour l'Europe, ici Paris.

ITALIE
457, 334, 225 en 190 m
(656, 899, 1331 en 1578 kHz)

18.00 Klankb. 18.45 Intern. Universiteit. 19.00 Dansmuz. 19.45 Caus. 20.00 Muz. Album. 20.30 Sport. 21.00 Gevar progr. 22.15 Licht ens. 22.30 Il conciliatore, van Samy Fayad. 23.15 Dansmuz. 24.00 Nieuws.

LUXEMBURG
1292 m (232 kHz)

8.40 Muz. v. d. huisvr. 11.43 Festival à Midi. 12.40 Dix millions d'auditeurs. 13.00 Nws. 13.15 Emportez-le avec vous! 13.22 Gevar. muz. 14.00 Tout autour du Monde. 14.10 Filmprogr. 14.25 Italiaanse muziek. 14.30 Bon Week-End! 14.50 A travers le monde. 15.00 Luxembourg Parade. 16.00 Protestantse act. 16.15 Radiokron 16.45 Luchtvaartkron. 17.00 Alors, raconte! 17.15 Celui que j'aime. 17.30 V. d. vrouw. 18.34 La vie comme elle va. 18.56 Belgische act. 19.00 Que faites-vous Dimanche? 19.05 Le Club de la Bonne Humeur. 19.15 Nieuws. 19.34 De toutes façons. 19.38 Dix millions d'auditeurs. 19.56 De familie Duraton. 20.05 Les risque-tout. 20.30 Le Point Commun. 20.46 Le rêve de votre vie. 21.15 Chansons. 21.40 Confidenties. 22.00 Le Bal Paris-Luxembourg. 23.55-24.00 Nws.

LUXEMBURG II
208 en 49.26 m
(1439 en 6090 kHz)

8.30 Muziek bij het huishouden. 10.30 Muziek bij een kopje koffie. 11.00 Verz. progr. 12.00 Lichte muz. 12.15 *Herman Emmink, de zingende omroeper.* 12.30 Sel. La Bohème, opera v. Puccini. 12.50 Nieuws. 13.04 Sportkalender. 13.15 Relax. 13.30-14.00 Prettig weekend!

ZWEDEN
1571, 505, 306, 254 en 245 m
(191, 593, 980, 1178 en 1223 kHz)
31.19 m (9.62 mHz)

19.00 Nws. 19.30 Gevar. progr. 20.30 Lichte muz. 21.00 Klankb. 21.30 Ork. o.l.v. Sten Frykberg, m.m.v. Göran Aakestedt, trompet: Werken v. Haydn. 21.50 Act. 22.15 Nieuws. 22.30 Dansmuz.

ZWITSERLAND
SOTTENS
393, 219, 49.55, 48.66 en 41.61 m
(764, 1367, 6055, 6165 en 7210 kHz)

19.15 Nieuws. 19.25 Tijdspiegel. 19.50 Cartes d'identité, door Gisèle Robert. 20.00 Incroyable, mais vrai!, door Claude Mossé. 20.35 Pension-Famille, hoorsp. 21.00 Discoparade. 22.00 Int. Gram. Concours te Brussel. 22.30 Nieuws. 22.35-23.15 Dansmuz.

Frequentie Modulatie

NORDDEUTSCHER RUNDFUNK (88.8 mHz)

17.30 Moderne amus. muz. 18.00 Der Sonntag und die gleitende Arbeitswoche, gesprek. 18.30 Strijkkwartet in cis kl. t., Pfitzner. 19.10 V. d. kind. 19.15 Muz. caus. 19.45 Uit het Middenduits dagboek. 19.55 Nieuws. 20.00 Vioolconc. in a kl. t., Bach. 20.20 Die Vergangenheit ist noch nicht beendet, fragm uit hoorspelen. 22.25 Dagboek v. d. Hong. revolutie. 22.30 Nieuws. 22.35 Sport. 22.50 Pianotrio op. 80 no. 2 in F gr. t., Schumann. 23.15 Geestelijke muz. 24.00-0.05 Progr. overz.

WESTDEUTSCHER RUNDFUNK
(95.7, 91.5, 90 en 89.7 mHz)

12.30 Nieuws. 12.45 Lichte muz. 14.00 V. d. kind. 14.30 Verz.-progr. 16.00 Die Eroberung des Nordpols, klankb. 16.40 Muz. discussie. 17.25 Keuls Omr. Symf. ork., m.m.v. Toos Onderdenwijngaard, piano. 18.30

Voordr. 19.00 Gewijde m
19.30 Tussen Rijn en
20.00 Nieuws. 20.10 Van d
dag. 20.15 Die Grosse S
rin, operette v. E. Kü
22.20 Harald Banter e
ens. 22.30 Sport. 23.00 N
23.05 Lichte muz. 24.00-0.
mermuz. v. Mozart en Me
sohn-Bartholdy.

Regionale Omro

NOORD, 188 m

19.30 Noordelijk weerb
Biologisch Kwartier. F
Brouwer beantwoordt b
19.45-20.00 Noordelijk
overzicht.

ZUID, 188 m

18.15 Nieuws. 18.20 De
lijkse felicitatierubriek „
leretribuun". 18.35 „Wa
Geer zoe wied van hoes
bezeuk bij Limburgers
diaspora. 19.00-19.30 „Spr
eure moond geit ope!" V
wedstrijd tösse gooj nao

WERELDOMROE

(Voor golflengten zie b
19.00 Nieuws. 19.10 Co
taar op het nieuws. 19.1
nieuwsjournaal, door Ge
Haan. 19.19 Ned ense
De Klompendansers o.l.
Steyn, met zang van
Brokken en Bert Robbe
Hersengymnastiek o.l.v
Boots en Jan van Herpe
Amusementsprogr. 20.46
gramma-overz. voor de
de week. 20.50 Nieuws.
Beursoverzicht. 20.55-21.
overdenking, door Pater
v. Croonenburg C.S.Sp.

K.R.O.-PUZZLE

HORIZONTAAL:

1 gedroogd vruchtvlees van de kokosnoot - gedwee - inwendig orgaan
2 planeet - elektrode
3 halfbloed - in overvloed neerdalen
4 loot - zangstem - bijwoord - ontkenning
5 deel van een schip - van bepaald soort hout - eenheid van spanning
6 wijd - dwaas - overleden
7 het radiobedrijf - constitutie
8 koopwaar
9 opgewekt - tulpachtige bloem
10 telwoord - waterdier - reclamelicht
11 paar - rechtsgeding - waardering
12 water in Utrecht - tijdsgewricht - bijwoord - metaal
13 gebruik - uitkering
14 nijver - plaats in Drente
15 slang - Engels bier - deel van een bruidstoilet

VERTICAAL:

1 vertrek - boom - bruine waterverf
2 geweldig - roofdier
3 Frans bacterioloog (1822-1895) - onbep. voornaamwoord
4 knaagdier - stad in Brazilië - ogenblik - herkauwer
5 lied - bedoelen - gast
6 deel van Sumatra - voorzetsel - buitgemaakt schip
7 beproeven - uitgestrekt plat land
8 nabootsing
9 steensoort - verbrandingsprodukt
10 ontkenning - schadelijke stof - teugel
11 gast - meisje - klipgeit
12 pl. in Gelderland - roofdier - plechtige belofte - vlezige uitwas
13 deelgenoot - toog
14 Nederlands vliegveld - intiem
15 appel - vrouw van Jacob - handigheid

Oplossingen kunnen worden ingezonden *uitsluitend op briefkaart onder vermelding „K.R.O.-puzzle".* U adresseert: K.R.O.-puzzle, postbus 1500, Amsterdam. De oplossingen dienen *vóór zaterdag, 26 okt. a.s.,* in ons bezit te zijn. Voor goede oplossingen staan tien prijzen ter beschikking.

(Voor oplossing en prijswinnaars zie blz. 15, 31, 36 en 59).

SPRINGPLANK-AUDITI

Deze week bestaat er weer ge
heid voor beoefenaars van de
muziek - zowel beroeps als ama
deel te nemen aan een auditie
het Springplankprogramma, e
in de volgende plaatsen:

Maastricht. Op dinsdag 22 o
in Hotel L'Universe, Kleinestr

Sittard. Op woensdag 23 okto
Hotel „De Zwaan", Brandstr

De audities vangen aan om 19.3
In de loop van het seizoen w
nog de volgende plaatsen be
Nijmegen, Winschoten, Leeuw
Leiden, Eindhoven, Breda, Ens
Apeldoorn, Middelburg en Delf

ONS ZENDERPAR

Hilversum I 402 m: 746

Hoogezand	FM 91.8	MHz	Kan
Hulsberg	FM 95.1	MHz	Kan
Markelo	FM 97.2	MHz	Kan
Irnsum	FM 96.0	MHz	Kan
Mierlo	FM 93.5	MHz	Kan

Hilversum II 298 m: 1007

Irnsum	FM 88.2	MHz	Kan
Hulsberg	FM 97.5	MHz	Kan
Markelo	FM 98.3	MHz	Kan
Hoogezand	FM 94.9	MHz	Kan

Van 7.00-10.00 en 17.00-24.00 u
golflengten 188 m: 1594 kc
Mierlo FM 93.5 MHz Kan

zaterdag 26 oktober ● KIJK OP BLADZIJDE 32 EN 33 ●

Fig. 2. Program guide text flow order and layout examples

percent of these terms are person names, and 17 percent are geographical names. A sample set was analysed and showed that 30 percent of these extracted terms match the terms that were manually added by archivists to the records in the NISV catalogue. This means that the remaining 70 percent adds new information to these records, mainly concerning the names of cast and crew.

IV. FUTURE WORK

The project was finalised by importing the matched records into the NISV catalogue. However, many catalogue systems, including the one of NISV are not well-suited yet for storing these links. For instance, it is not possible to click on through to a program guide from the record of matched TV program. Catalogue systems would require significant changes in order to truly facilitate storing contextual links and for machines to access them through APIs. Also, applications need to be developed that allow end users to fully explore these links.

Furthermore, the parsing script, configuration files and matching scripts can be improved to increase accuracy of the output in various stages of the workflow. This is not just needed to improve results for the set we have used for this pilot contextualisation project, but especially when scaling up. Besides the 330 guides used for the pilot, the archive has scanned 3,300 more guides from six Dutch public broadcasters from 1951-2014, covering 250,000 pages containing the complete overview of what was broadcast on Dutch television and radio.

Finally, we would like to explore the contextualisation possibilities further, by creating links between the program guides and the archive's large photo collections and (linked) data initiatives like Wikidata[8] to come full circle and connect not only our internal collections, but to link to and from relevant external sources as well.

ACKNOWLEDGMENTS

This work was made possible by the large-scale digitisation programme Images for the Future (2007-2014). The parsing script was developed in collaboration with Thijs Leegwater from Picturae, to whom the authors are grateful.

REFERENCES

[1] G. Auffret and B. Bachimont, "Audiovisual cultural heritage: from tv and radio archiving to hypermedia publishing," in *Research and Advanced Technology for Digital Libraries*. Springer, 1999, pp. 58–75.

[2] D. W. Oard, A. S. Levi, R. L. Punzalan, and R. Warren, "Bridging communities of practice: Emerging technologies for content-centered linking," in *Museums and the Web*, 2014.

[3] C. A. Harper, "Linked open data in libraries, archives and museums (lodlam): current trends, tools & techniques, and the role of vendors," 2013.

[4] M. Bron, J. van Gorp, F. Nack, M. de Rijke, A. Vishneuski, and S. de Leeuw, "A subjunctive exploratory search interface to support media studies researchers," in *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ser. SIGIR '12. New York, NY, USA: ACM, 2012, pp. 425–434. [Online]. Available: http://doi.acm.org/10.1145/2348283.2348342

[8] http://www.wikidata.org/