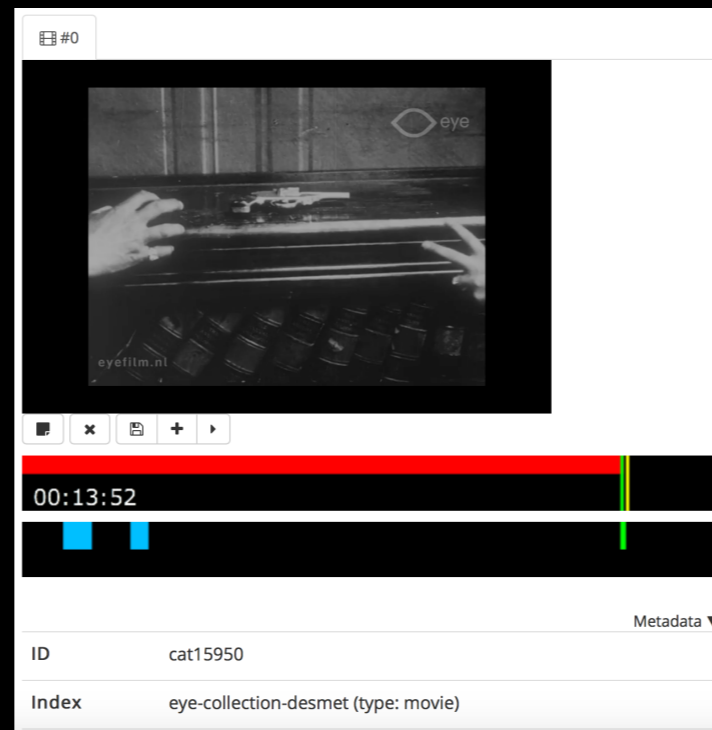


MIMEHIST: Annotating Eye's Jean Desmet Collection



AMIA, Portland, November 29, 2018



What we do in MIMEHIST



Jean Desmet (1875-1956)

- CLARIAH Mediasuite Pilot Project: UvA, EYE Filmmuseum, NISV
- Desmet Collection and EYE's "Defragmentation" problem
 - Video annotation as a solution.
- Allow for linking and annotating materials in personal user spaces
 - Embed Desmet Collection in the CLARIAH Media Suite
 - Automated analysis of business archive

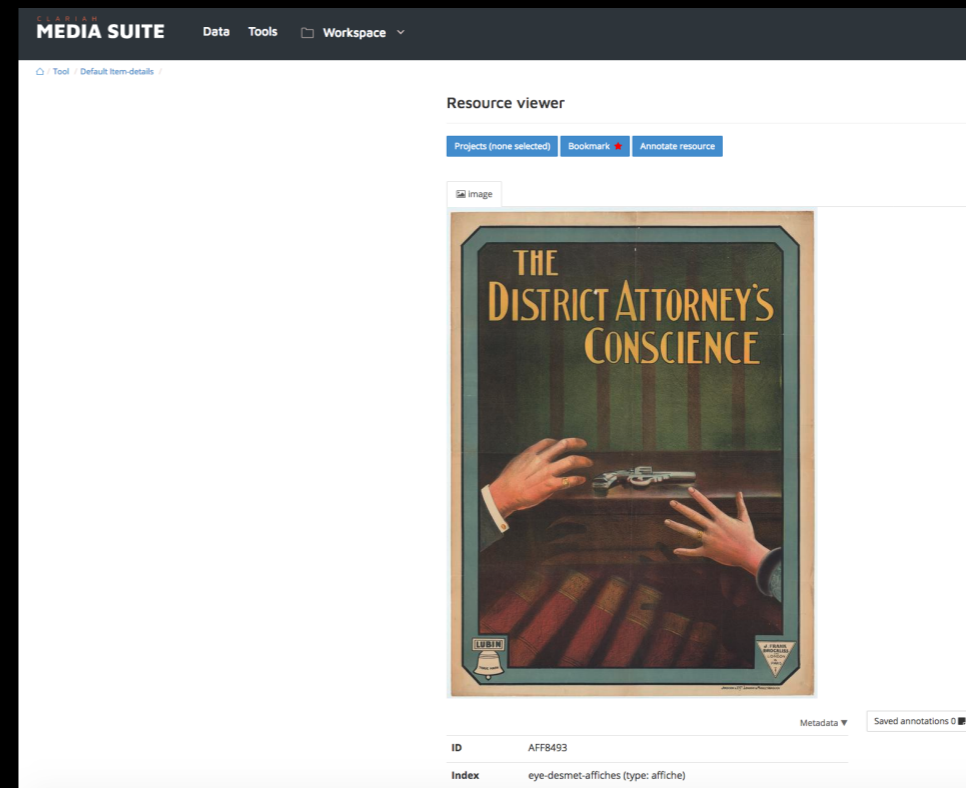
Bookmarking and Corpus Building

The screenshot displays the Media Suite interface for a workspace named 'District Attorney's Conscience'. The interface includes a navigation bar with 'Tools', 'Workspace', 'About', and 'Documentation'. Below the workspace name, there are tabs for 'Details', 'Bookmarks' (72), 'Annotations' (0), 'Tool sessions' (0), and 'Queries' (0). A search bar and filter options are present, including 'Search Bookmarks', 'Media', 'Group', and 'Annotations'. A table lists the bookmarked items with columns for 'TITLE', 'DATE', 'MEDIA', 'DATASET', and 'GROUPS'. Each row includes a 'VIEW' button and a '12' icon.

TITLE	DATE	MEDIA	DATASET	GROUPS
District Attorney's Conscience, The (original title)	1913	video	eye-desmet-films	FILM
Fixing a Flirt (original title)	1912	video	eye-desmet-films	FILM
Work in a United States Arsenal Work in an US Arsenal (original title)	1912		eye-desmet-films	FILM
Oud en jong (Dutch title)	1910 - 191		eye-desmet-films	FILM
Literature and Love (original title)	1913	video	eye-desmet-films	FILM

- Corpus building in personal user spaces
- EYE's complete Desmet Collection Metadata
- Films, posters and business archive embedded via EYE server

Segmentation and Annotation



The District Attorney's Conscience (Lubin, US, 1912)

- Going beyond Eye's archival finding aid
- Mediasuite annotation functionalities for linking films and related documents
- Explore a film's distribution and exhibition history through e.g. posters, letters

Automated analysis of business archive



- Materially extremely heterogeneous business archive
- EYE's metadata is an excellent starting point, but only folder-specific
- Unlock business archive with OCR and AI methods
- Data included in the CLARIAH Mediasuite with EYE's Metadata

Handwriting identification

ferdinand

Sittard 28. 1. 1913

waarde Broer en familie

desen Bericht ik u als dat
wij voorlopich de bioscoop gists
hebben denkt dat u

Plus ma g
laat afstuur
saken Sche
weet dat
Carnaval over

vertels gy
Gr
en

henriette

daar he atijts rich terug tekt en Ferdinand
rijn adres niet weet. en hierover ontfang
ik van hem ook een teken gesteld ant
woord. Maar enjin de tijden zullen er
maken een teken. Nu heb ik natuur
lijk dat huis moeder of geven en te beken
hijns tijt de huizen een leugen moeten
meesmaker, Soan er natuur tijt, too
veel waaning word it seft bij
die contract breuk niet too
afgeven Wel jammer ik haalden
daar mel eene batenham gevonden
en teu ik mel meer met wat
ander in gesprek maan niet too
gelegen en niet too groot. Maar ik
wilt nog wat afwaachten van, In dien
het milte wordt too bericht ik te.
en moeden ik wat anders hebben rult
u wel hooren van mijn Eld u Ferdinand
rijn adres weet wil je mijn dan
even Bericht geven

Gruet van Charles Catherine, en Jeanne
gehuut van mijn en Jette

Ume bijden mijn bestes
graeten

Henriette

Jette is war beter

theo

leveranties van films, zoo als hij mij schreef zoudt
hij dat in de films verhuurder, vergadering
voorstellen; en het puntje bepaald zoudt
komen om hem dan te voldoen, moey ik
aan Frank spoedig besluit van het, al of niet
geuen van het geb, m
gram zou bijden m
hem herken het gona
afdaen bij te
uer moest hi
mij dat hij ha
te een telegram
den

Waarde Bened
Hep is zoo als
in met Pij, Ma
sh.
ieromkent gel d
opbellen. Nu
hoog mel een. n
t dan mel ver
o van hem nem
want moest v
rent klaar at
njin der dard
kon mij daan
9

ella

ik nog gehoord dat ry de vliegen
erg aantrekt en daarom heel veel
op de muren van veestallen ge-
bruikt wordt. Herklig heb ik mijn
all in mijn keuken over de vliegen
drukte latsen zover gevonden en
waren heel gauw van de aklige stips
te rieu.

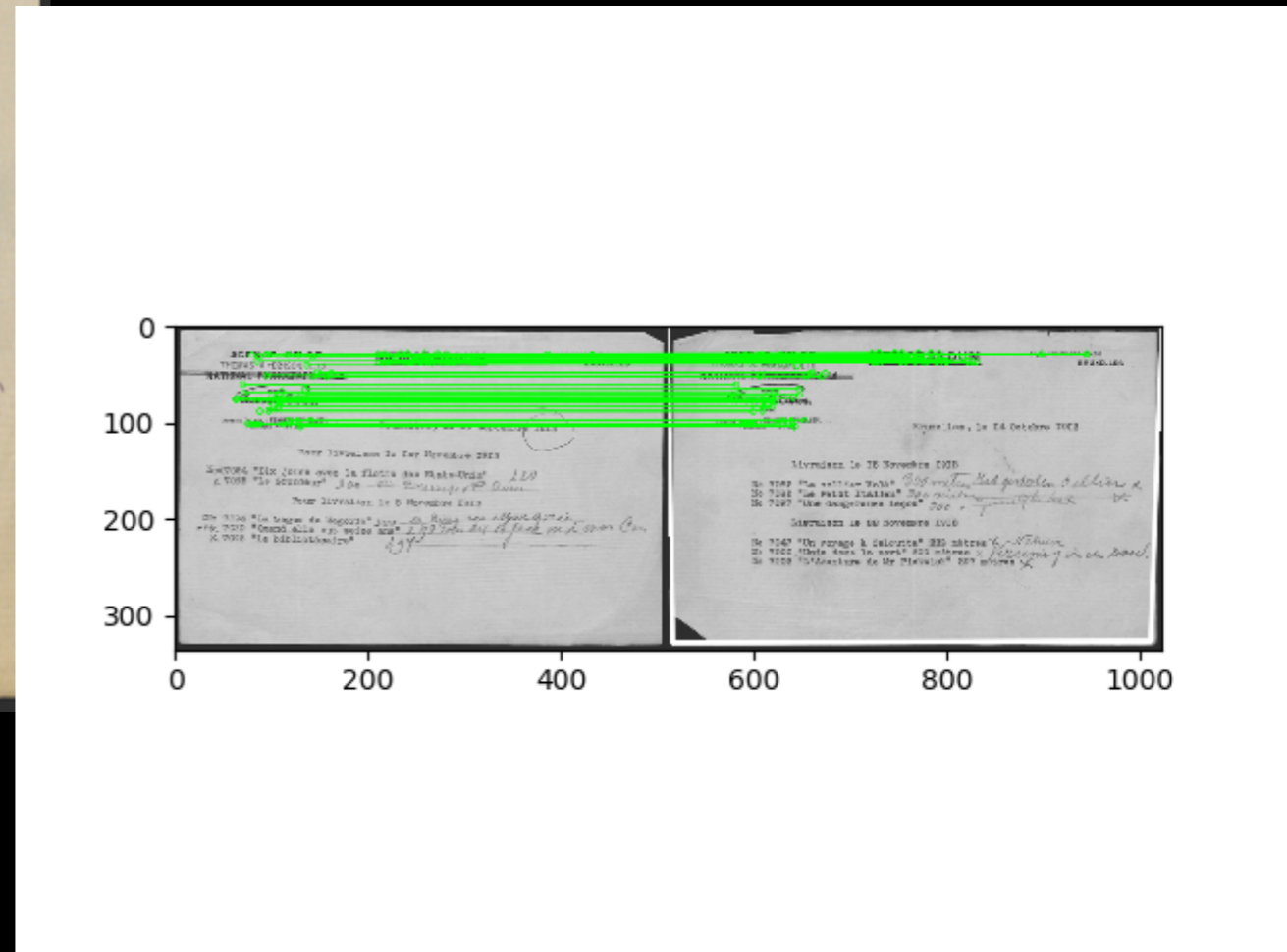
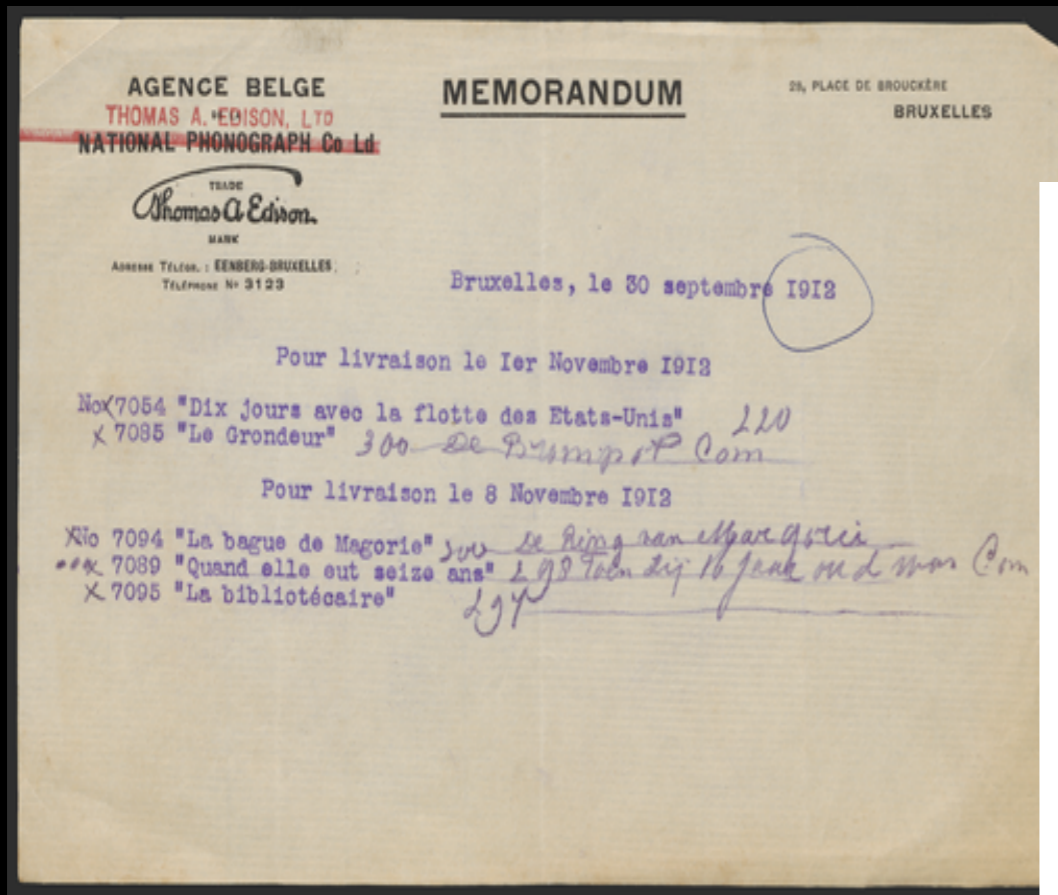
hoe vind je de indeling der
keuken op dat ball. plaatje? For een
reer groote k. mel een gerieflijk vagenke

Mu, Zica, ik sluit desen
brief, Thys doet hem direkt op de
post. veel succes met de nieuwe
woninginrichting!

Hartelyke graeten ook van
Thys voor jocht en Jan en Jeanne
van je schoonzuster

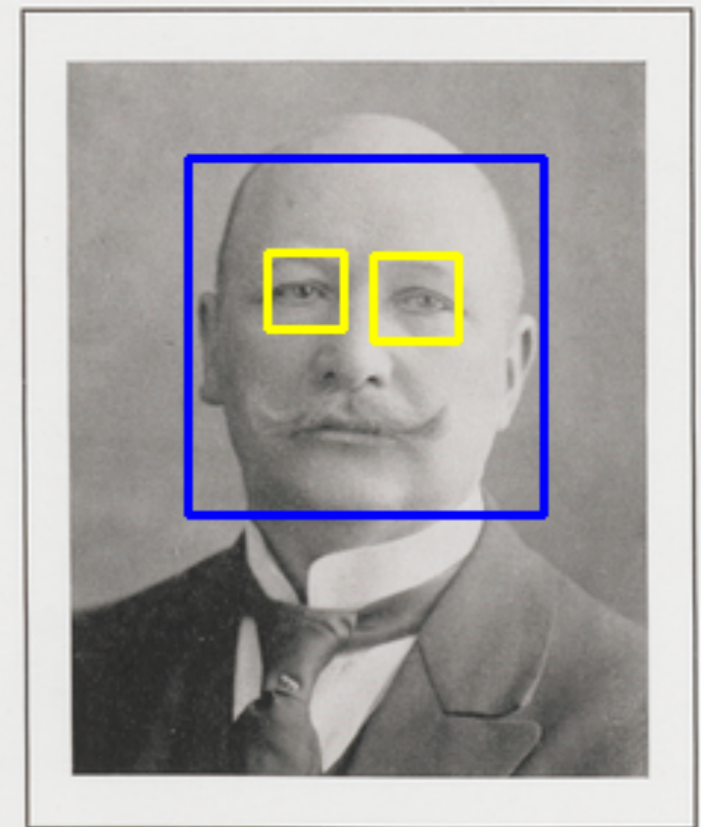
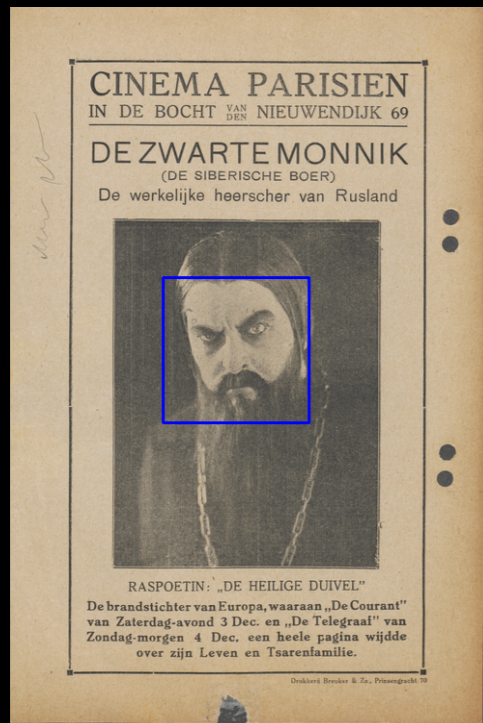
Ella.

Similar layouts



- Similarity index comparison + SIFT (scale-invariant feature transform)
- 27758 documents into 815 clusters,
 - particular individuals, specific institutions, cinemas, production companies, government institutions, grocery stores, electricians, tax forms, rental contracts, film lists, etc.

Face detection



JEAN DESMET

- Frontal face
- Face profile
- Eyes
- Upper body

Business Documents in Media Suite

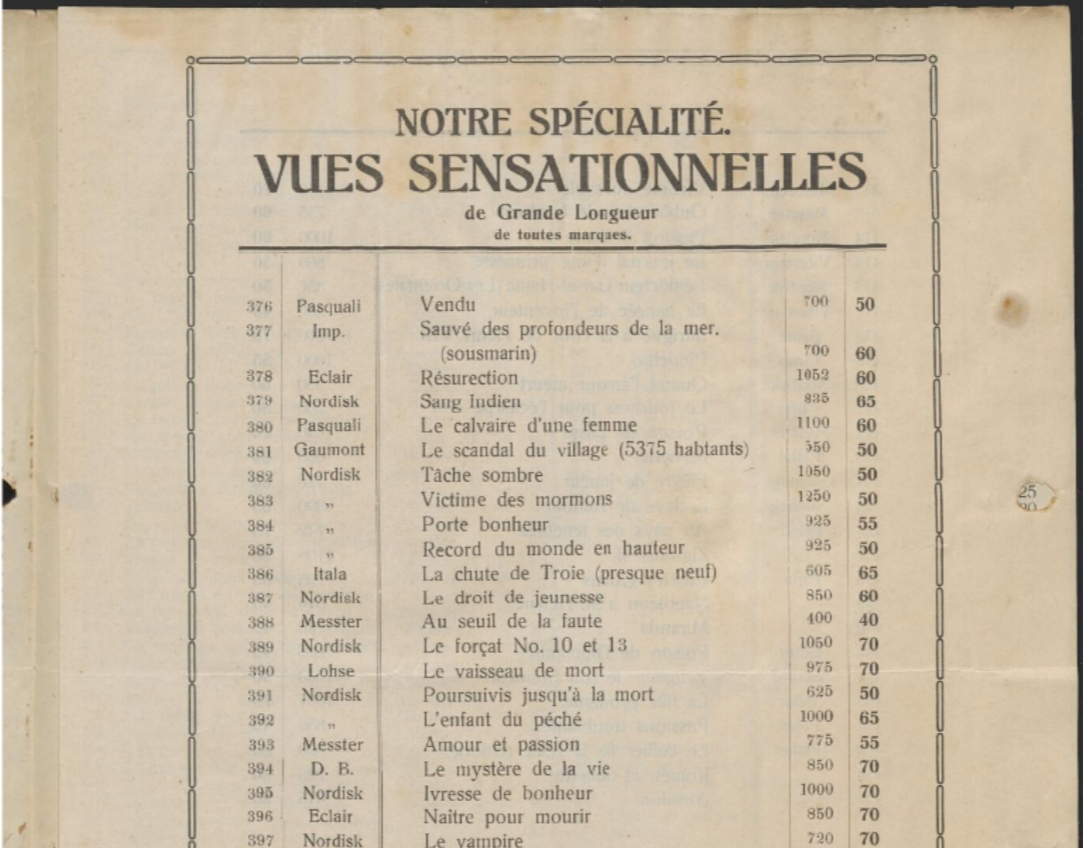
Tools Workspace About Documentation

Resource viewer

Projects (Dødsspring til Hest fra Cirkuskuplen) Bookmark ★ Annotate resource

Related media objects ▼

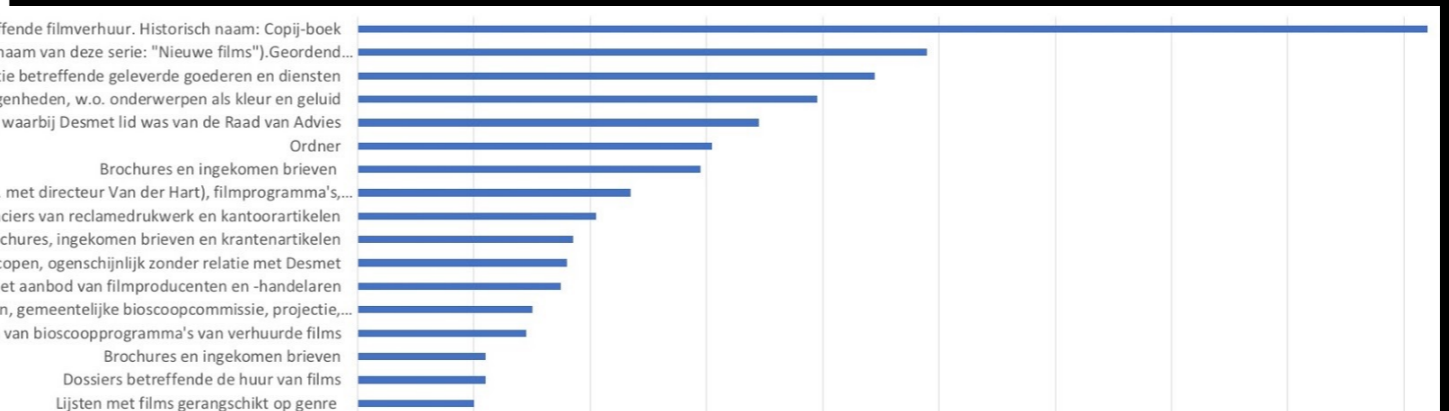
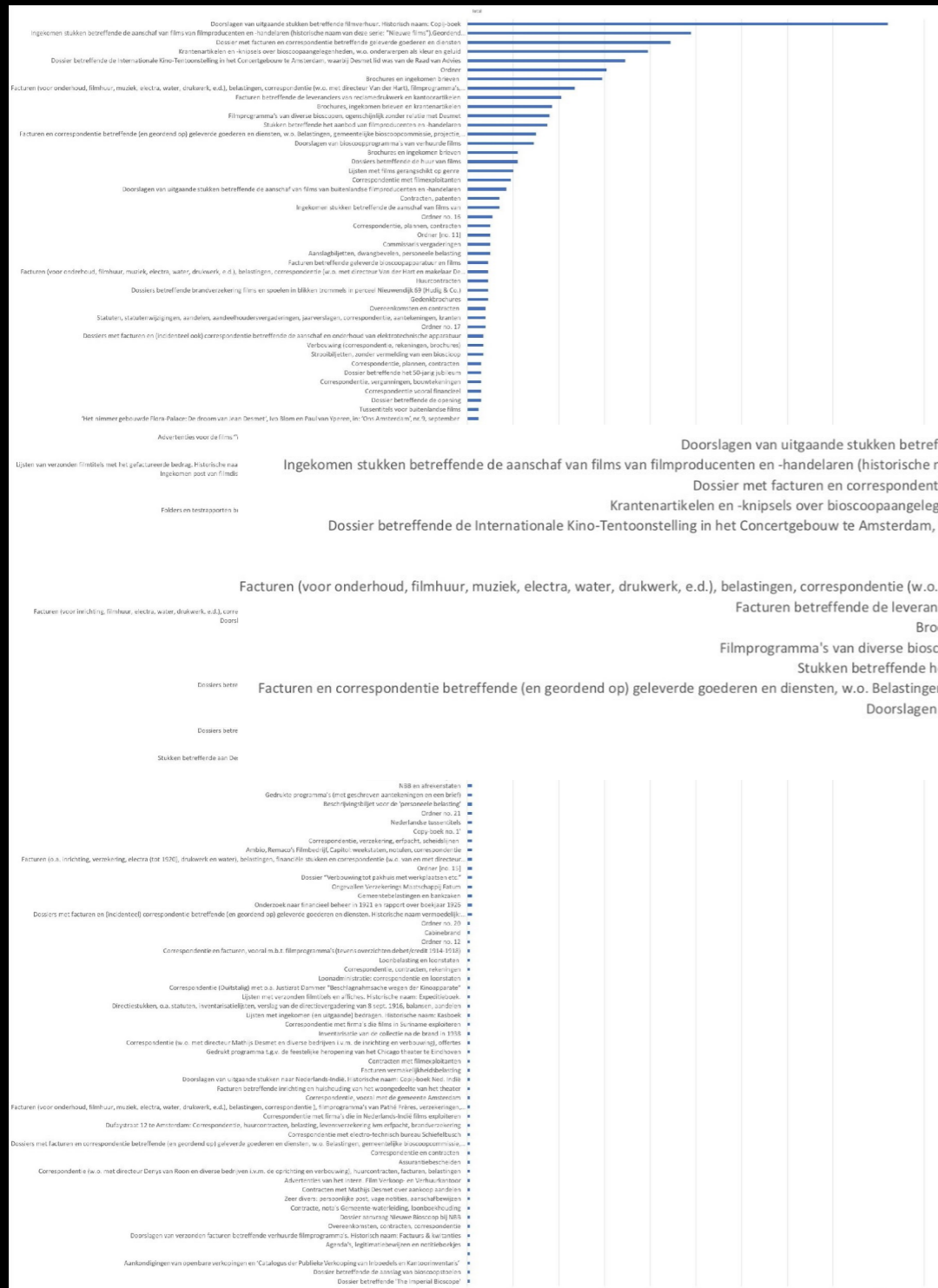
image



376	Pasquali	Vendu	700	50
377	Imp.	Sauvé des profondeurs de la mer. (sousmarin)	700	60
378	Eclair	Résurrection	1052	60
379	Nordisk	Sang Indien	895	65
380	Pasquali	Le calvaire d'une femme	1100	60
381	Gaumont	Le scandale du village (5375 habitants)	350	50
382	Nordisk	Tâche sombre	1050	50
383	"	Victime des mormons	1250	50
384	"	Porte bonheur	925	55
385	"	Record du monde en hauteur	925	50
386	Itala	La chute de Troie (presque neuf)	605	65
387	Nordisk	Le droit de jeunesse	850	60
388	Messter	Au seuil de la faute	400	40
389	Nordisk	Le forçat No. 10 et 13	1050	70
390	Lohse	Le vaisseau de mort	975	70
391	Nordisk	Poursuivis jusqu'à la mort	625	50
392	"	L'enfant du péché	1000	65
393	Messter	Amour et passion	775	55
394	D. B.	Le mystère de la vie	850	70
395	Nordisk	Ivresse de bonheur	1000	70
396	Eclair	Naitre pour mourir	850	70
397	Nordisk	Le vampire	720	70

- 127.000 business documents in Media Suite, OCR data added

Business archive corpus building 1



- Only one of these folders has one of the keywords in its name
- Type-written sources come from 136 different folders (largest category: 184)

Test Cases



- Test Case 1: *Dødsspring til Hest fra Cirkuskuplen* (Dir.: Schnedler-Sørensen, Prod.: Nordisk Films Kompagni, 1912)
 - create a “historical-critical” edition primarily with what is available in the Media Suite

MIMEHIST results

- A significant and widely used historical collection added to the CLARIAH infrastructure
- Unlocking Desmet's business archive and gain new insights into distribution/exhibition history with data enrichments
- Experiment with combining methods: OCR *and* visual classification can support each other

Thanks for your attention!

c.g.olesen@uva.nl