



**ifis**

Institut für Informationssysteme  
Technische Universität Braunschweig

KB } national library  
of the netherlands

# **Aspect-Driven Structuring of Historical Dutch Newspaper Archives**

**Hermann Kroll, Christin K. Kreutz, Mirjam Cuper  
**Bill Matthias Thang and Wolf-Tilo Balke****

Institute for Information Systems  
TU Braunschweig, Germany



# Why we are here?

**TPDL 2022**  
The 26th International Conference on Theory and Practice of Digital Libraries

**Location**  
Padua, Italy

**Dates**  
0 - 23 September 2022

**Deadlines**

Full:	May 29, 2022
Short:	June 5, 2022
Innovation:	June 17, 2022
Workshops:	Feb. 14, 2022
Doc. Cons.:	1 August 2022

TPDL 2022 aims to be an in-presence event.

[Follow @tpdl2022](#)

[Log in @tpdl2022](#)

**You are here**



**Hermann**



**Mirjam**

KB } national library  
of the netherlands

# **The National Library of the Netherlands**







# The National Library of the Netherlands

Delpher

Thema'sFavorietenInstellingenHandleidingOver Delpher

KrantenolifantZoekenUitgebreid zoeken

Terug naar kranten

Periode

☐ 17e eeuw (42)

☐ 18e eeuw (252)

☐ 19e eeuw (15126)

☐ 20e eeuw (104517)

☐ 21e eeuw (553)

Kies periode...

Soort bericht

☐ Advertentie (23184)

☐ Artikel (91365)

☐ Familiebericht (175)

☐ Illustratie met onderschrift (5766)

Verspreidingsgebied

☐ Landelijk (42935)

☐ Nederlands-Indië / Indonesië (14100)

☐ Nederlandse Antillen (937)

☐ Regionaal/lokaal (61071)

☐ Suriname (696)

☐ Verenigde Staten (751)

Krantentitel

Kies krantentitel...

Plaats van uitgave

Kies plaats van uitgave...

Herkomst


Kies herkomst...

120.490 krantenartikelen gevonden

ZoektermOlifant

Sorteer opdatum [0-9]

Weergave




NEDERLANDEN:

ettfe Dagen heeft fijn Hoogheyr donHeere 'Prince van Orangie fijnen eerften Cam-r'Ditnaer, den Heer **Olifant** / dér Rechten Doftor, pebenliceert me" de Balliws-plaette van Wil' Iem Stadt/

KrantentitelOprechte Haerlemsche courant

Datum24-12-1669

Meer details




NEDERLANDEN.

elevertover 't Arbitragie-werck. Amsterdam den 20 Maert. 't Schip Broeder, Schipper Ian Cornelisz. **Olifant**, is niet genomen , gelijk men over Marsielje heeft Tydinge gehad; maer is met noch een Geno

KrantentitelOprechte Haerlemsche courant

Datum21-03-1671

Meer details




NEDERLANDEN.

m den 17 Mey. De grooteSchepenvan Oorlogezijn alle over 't Pampus geraeckt, uyngenomen 't Schip den **Olifant**. De Fransse beginnen haer omtrent de nieuwe Post tot Nieuwer-sluyts te vertoonen, soo dat men

KrantentitelOprechte Haerlemsche courant

Datum18-05-1673

Meer details




NEDERLANDEN.

hepen van Oorloge Zijn-<£, alle over't Pampus v/raeckt, uyngenomen't Schip den **Olifant**.' De Franfle beginnen haer omtrent de nieuwe Pof tot NieuWer-luys te Vmoonen, foo dat men

KrantentitelOprechte Haerlemsche courant

Datum18-05-1673

Meer details



Duytsland en d'aengrenfende Rycken.

De Heer Buchwalt is na Coppenhage verreyt , zijnde van sijn Majesteyt vereert met d'Ordre van den **Olifant**.

Aspect-Driven News Structuring — Mirjam Cuper / Hermann Kroll — KB / TU Braunschweig

4






# Research Question

How can a digital library design effective access paths to explore their collection?

## De Watersnoodramp van 1953



### Berichtgeving over de stormvloed

De Watersnoodramp begint zeventig jaar geleden, in de nacht van zaterdag 31 januari op zondag 1 februari 1953. Een zware noordwesterstorm in combinatie met springtij, zorgt ervoor dat veel - slecht onderhouden - dijken doorbreken. Grote delen van Nederland overstromen. Vooral Zeeland en de Zuid-Hollandse eilanden worden getroffen. De hulpverlening komt traag op gang en er vallen veel slachtoffers; in ons land komen 1.836 mensen om. Ook de materiële schade is enorm.

De ramp heeft grote gevolgen: voor de overlevenden, die decennia later nog kampen met trauma's, maar ook voor de kustverdediging van Nederland.

Volg het verloop van de ramp en de nasleep ervan aan de hand van krantenberichten van toen. Van de eerste dijkdoorbraken tot de bouw van de Deltawerken. Wil je meer lezen over de Watersnoodramp? Ga dan vooral ook zelf op zoek in Delpher; er is nog veel meer over dit onderwerp te vinden.

### Eerste berichtgeving

#### Noodtoestand

**ANP-radiobulletin, 1 februari 1953**

Op zondag 1 februari verschijnen er geen kranten. En dus zijn de krantenredacties verlaten als in de nacht van zaterdag op zondag de eerste berichten over dijkdoorbraken binnenkomen. Alleen bij de Radionijsdienst van het ANP wordt gewerkt. Om 8.00 uur 's ochtends vertelt de nieuwslezer over "een noodtoestand" die is ontstaan door "abnormaal hoge waterstanden". De omvang van de ramp is op dat moment nog lang niet bekend. Het rampgebied is vrijwel onbereikbaar; wegen en spoorwegen staan onder water; telefoonverbindingen zijn verbroken. Hoewel het ANP meldt dat "het hoogtepunt van de zeer zware storm" ons land gepasseerd is, moet het ergste in werkelijkheid nog komen.

[Bekijk de tekst van het radiobulletin](#)

#### Allesverwoestend

**Algemeen Dagblad, 2 februari 1953**

De kranten staan op 2 februari vol met berichten over de stormvloed. "Hoewel de toestand nog zeer onoverzichtelijk is en van overal de berichten over dijkdoorbraken blijven binnenstromen - later soms weer tegengesproken", probeert Het Algemeen Dagblad orde in de chaos te scheppen. De krant geeft een opsomming van de laatste berichtgeving en voegt ter verduidelijking een kaartje toe. De Koninklijke Nederlandse Automobil Club (KNAC) doet ondertussen een beroep op automobilisten "om zich waar dit van nut kan zijn ter beschikking te stellen van de hulporganisaties."

[Bekijk deze krant](#)

#### Niets dan water

**Het Parool, 2 februari 1953**

Langzaam wordt duidelijk hoe ernstig de situatie is. In de middag van 1 februari bereikt een tweede vloed het rampgebied. Deze eist nog meer slachtoffers dan de eerste. Een "verkeersredacteur" van Het Parool maakt op 2 februari een vlucht boven de getroffen gebieden en meldt dat Goeree-Overflakkee en Schouwen-Duiveland zijn verdwenen. "Wie boven deze eilanden vliegt ziet werkelijk alleen nog maar water en hier en daar een dak van een huis." Vanuit de lucht ziet het er vredig uit, maar "hier hebben zich drama's afgespeeld, hier hebben mensen die ten onder gingen elkaar voor het laatst in de ogen gekeken."

[Bekijk deze krant](#)

#### Dodental loopt op

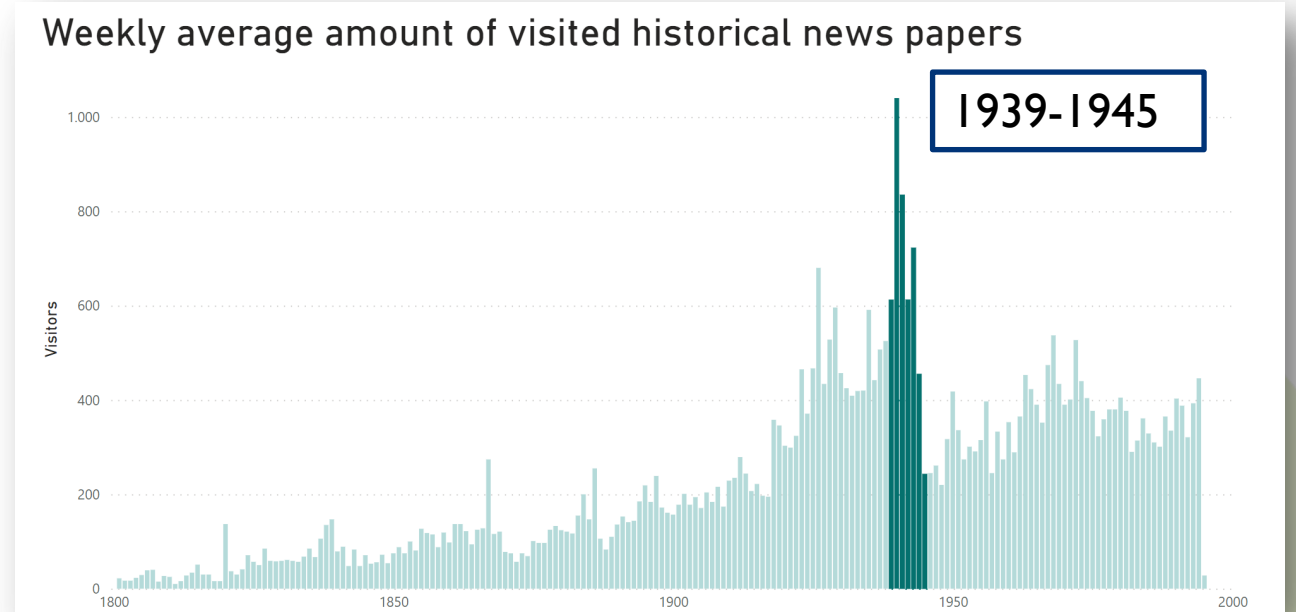
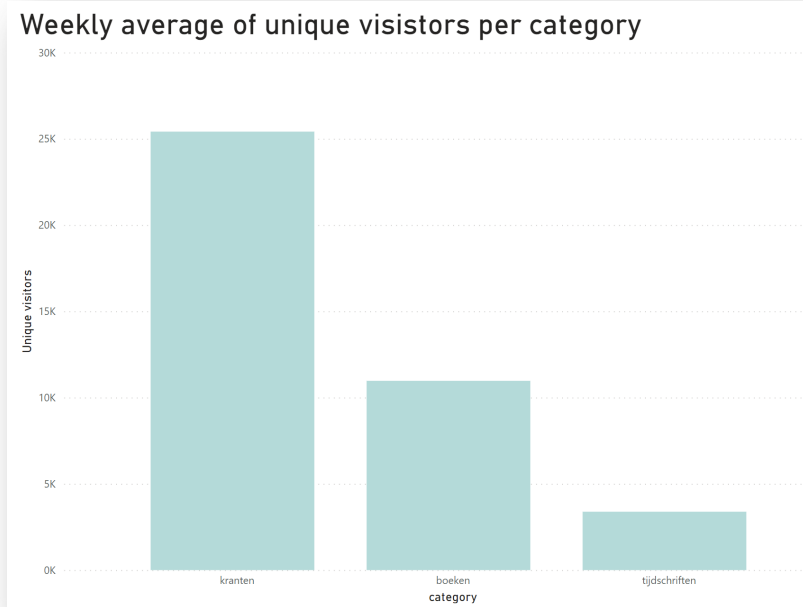
**De Waarheid, 2 februari 1953**

De "droeve lijst" op de voorpagina van De Waarheid van 2 februari laat zien dat het dodental snel oploopt. Al is op dat moment nog "van talloze landgenoten onbekend of zij in veiligheid zijn dan wel eveneens door het water zijn versloegen". De communistische krant, altijd kritisch op het overheidsbeleid, wijst ook op de slecht georganiseerde hulpverlening. Die komt inderdaad moeizaam op gang. Het overzicht ontbreekt, bovendien heeft Nederland maar één helikopter. Burgers proberen elkaar zo goed en zo kwaad als het gaat te helpen.

[Bekijk deze krant](#)



# Constraints



- **Dutch** data, but NLP methods often in English
- News articles based on **OCR** (with typical OCR issues)
- **License** prohibiting sending data to APIs like ChatGPT



**ifis**

Institut für Informationssysteme  
Technische Universität Braunschweig

# **An Aspect-Driven Access Path**





# Basic Concepts

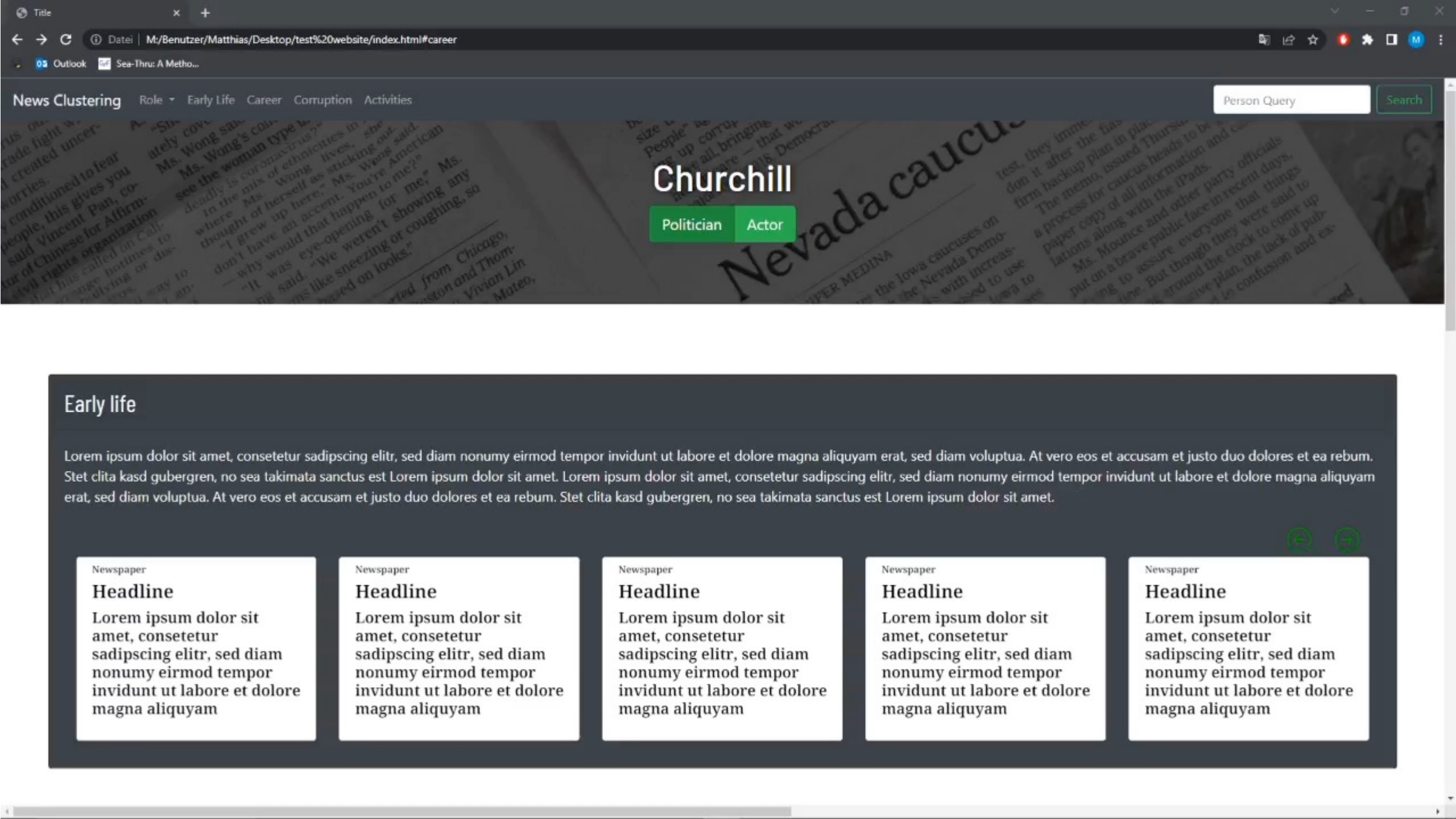
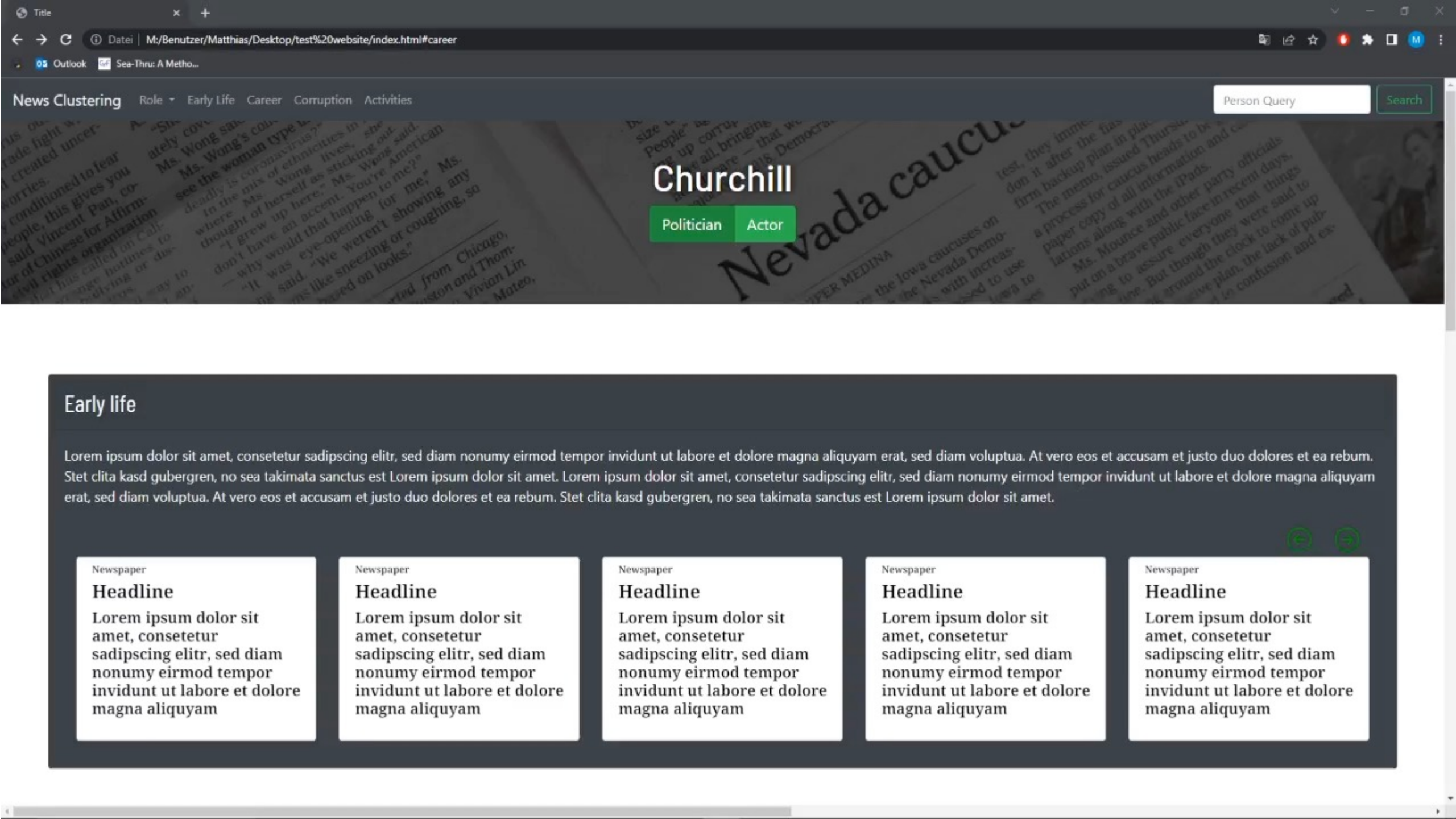
- **News article** has:

- Unique identifier
- List of paragraphs



- **Person** is identified by a name plus titles (synonyms):

- **Person Class:** job/position, e.g., politician, artist, writer, scientist, ...
- **Person Aspect:** “Early Life”, “Political Career”, ...



# Dwight Eisenhower

minister

persoon

politicus

achtergrond | afkomst en opleiding | biografie | biografische schets | jeugd | jeugd en opleiding | jeugd jaren | jonge jaren | leven | leven en werk | levensloop | vroege jaren | vroege leven

Dwight Eisenhower is geboren op 14 oktober 1890 in Denison, Texas. Zijn familie verhuisde naar Abilene in Kansas een paar maanden na zijn geboorte, waar hij en zijn broers "op een typisch landelijke manier" werden opgevoed, volgens de Wikipedia pagina. Hij studeerde af aan de Columbia University en promoveerde aan de Universiteit van Chicago. Hij diende in de Tweede Wereldoorlog als adjunct-hoofd van psychologische oorlogvoering op het hoofdkwartier van de geallieerde in Noord-Afrika, en later dezelfde functie in Europa, volgens Wikipedia. Hij was de vader van president Harry S. Truman, minister van Defensie Donald Rumsfeld en minister van Buitenlandse Zaken John F. Kennedy. Hij had ook een zus, Barbara Eisenhower, die "geen of fictieve functie heeft op de reis van de president," schrijft Wikipedia. Eisenhower was de zoon van "onbekende, onopvallende en hardwerkende" ouders, de site notities. Hij had vijf broers en een zus, "zus," die met haar familie in 1871 naar de VS kwam, de website notities. Zijn broer John Foster Dulles was Eisenhower's minister van Buitenlandse Zaken en zijn schoonvader was de eerste Amerikaanse ambassadeur in Denemarken. Hij stierf in 1953. Klik voor de volledige lijst van Eisenhower familieleden.

Algemeen Handelsblad (1964-09-28)

DE COMMISSIE-

De nieuwe Limburger :  
provinciaal dagblad [Editie  
Maastricht en omgeving] (1960-02-10)

Nieuwe Tilburgsche courant:  
dagblad voor centraal Brabant :  
Tilburg dagblad (1956-11-08)

De leenbaan van

Nieuwsblad van het Zuiden :  
dagblad met ochtend- en avond-  
editie (1960-02-06)

Negentigjarige

De nieuwe Limburger :  
provinciaal dagblad [Editie  
Maastricht en omgeving] (1956-11-08)



## loopbaan

De New York Times blikt terug op president Eisenhower... de eerste volle dag in functie, die begon met de benoeming van veldmaarschalk Burggraaf Montgomery als zijn plaatsvervanger. Hier zijn enkele van de opmerkelijke afspraken gemaakt die dag: Staatssecretaris John F. Kennedy als zijn minister van Buitenlandse Zaken. Minister van Financiën William H. Hobby als zijn hoofd van de Gezondheidszorg Toezicht Office. Generaal Eisenhower als zijn persoonlijke vertegenwoordiger bij de NAVO. De voorzitter van de Commissie buitenlandse zaken van de Senaat, John J. Connally, en de voorzitter van de House Committee on Commerce, Science, and Transportation, werden benoemd als de plaatsvervangers en commandanten van Eisenhower. De Republikeinse Partij-ambtenaar Steve Anderson nam ontslag uit protest tegen Eisenhower's steun aan senatoren McCarthy en Jenner, die tegen het communisme waren. De Republikeinse kandidaat voor president Paul G. Hoffman verklaarde dat Eisenhower president kon worden als Republikein zonder kandidaat te worden, de Wall Street Journal rapporten. Eisenhower benoemde voormalig Minnesota Gov. Harold Stassen als zijn directeur van de Amerikaanse Geheime Dienst. Hij benoemde de voormalige Amerikaanse ambassadeur in de Sovjet-Unie, Herbert Brownell Jr., als zijn procureur-generaal. Eisenhower promoveerde Adm. William J. Burke tot admiraal, bijgenaamd "Justin Knots" omdat hij tijdens de Tweede Wereldoorlog liever torpedobootjagers dan torpedobootjagers. Hij benoemde de voormalige Amerikaanse onderminister van Buitenlandse Zaken John B. Stikker tot zijn speciale gezant in Rusland. De Republikeinse Nationale Commissie werd hernoemd tot Republikeinse Partij van de Verenigde Staten van Amerika als reactie op een brief van Eisenhower aan de National Republikeinse Senatorial Committee. In zijn meest wonderlijke vorm, komt het nu neer op bier, die Eisenhower moet blijven regeren met de Who of the Democrats, citeert de Daily Mirror Hoffman als gezegde.

Trouw (1952-11-22)

**Nieuwe ministers  
van Eisenhower:  
Stassen, Humphrey,  
Brownell  
„Onderzoek mij”**

Fragment:  
[...]tot directeur van de  
MSA. Brownell. advocaat te

Amigoe di Curaçao : weekblad  
voor de Curaçaosche eilanden  
(1948-07-03)

**Britten eisen  
opheffing blokkade  
van Berlijn Zoekt  
Joegoslavië  
toenadering tot de  
Verenigde Staten**

Nieuwsblad van het Noorden  
(1953-03-27)

**Bohlen en Moskou**

Fragment:  
[...]Roosevelt (democraat)  
en Truman (democraat).  
Thans staan de  
republikeinen Eisenhower  
en Dulles borg voor zijn  
integriteit. Bohlen en

Algemeen dagblad (1959-05-25)

**JOHN FOSTER  
DULLES - DE  
VECHTER**

Fragment:  
[...]specialist op het gebied  
van buitenlandse politiek,  
en toen generaal  
Eisenhower in 1953 het

Provinciale Drentsche en Asser  
courant (1951-03-21)

**Eisenhower heeft  
zijn commandanten  
uitgekozen  
Montgomery zijn  
voornaamste  
assistent**

Fragment:

# Dwight Eisenhower

minister

persoon

politicus

Kies een Persoon ▼

Winston Churchill

Dwight Eisenhower

Simon Vestdijk

Leopold III

Frans Goedhart

Franklin Roosevelt

Wilhelmina van Oranje-Nassau



Algemeen Handelsblad (1964-09-28)

## **DE COMMISSIE- WARREN**

Fragment:

[...]hij hoofd van de  
Centrale

Inlichtingendienst. Zijn  
broer John Foster Dulles  
was Eisenhowers minister  
van buitenlandse zaken.

Allen woont nu als  
advocaat in[...]

De nieuwe Limburger :  
provinciaal dagblad [Editie  
Maastricht en omgeving] (1960-  
02-10)

## **Gemengd nieuws**

Fragment:

[...]vergezeld door haar  
familie en de zuster van  
haar vader die met een  
Eisenhower was getrouwd.  
Deze laatsten waren de  
ouders van president  
Eisenhower.[...]

Nieuwe Tilburgsche courant:  
dagblad voor centraal Brabant :  
Tilburg dagblad (1956-11-08)

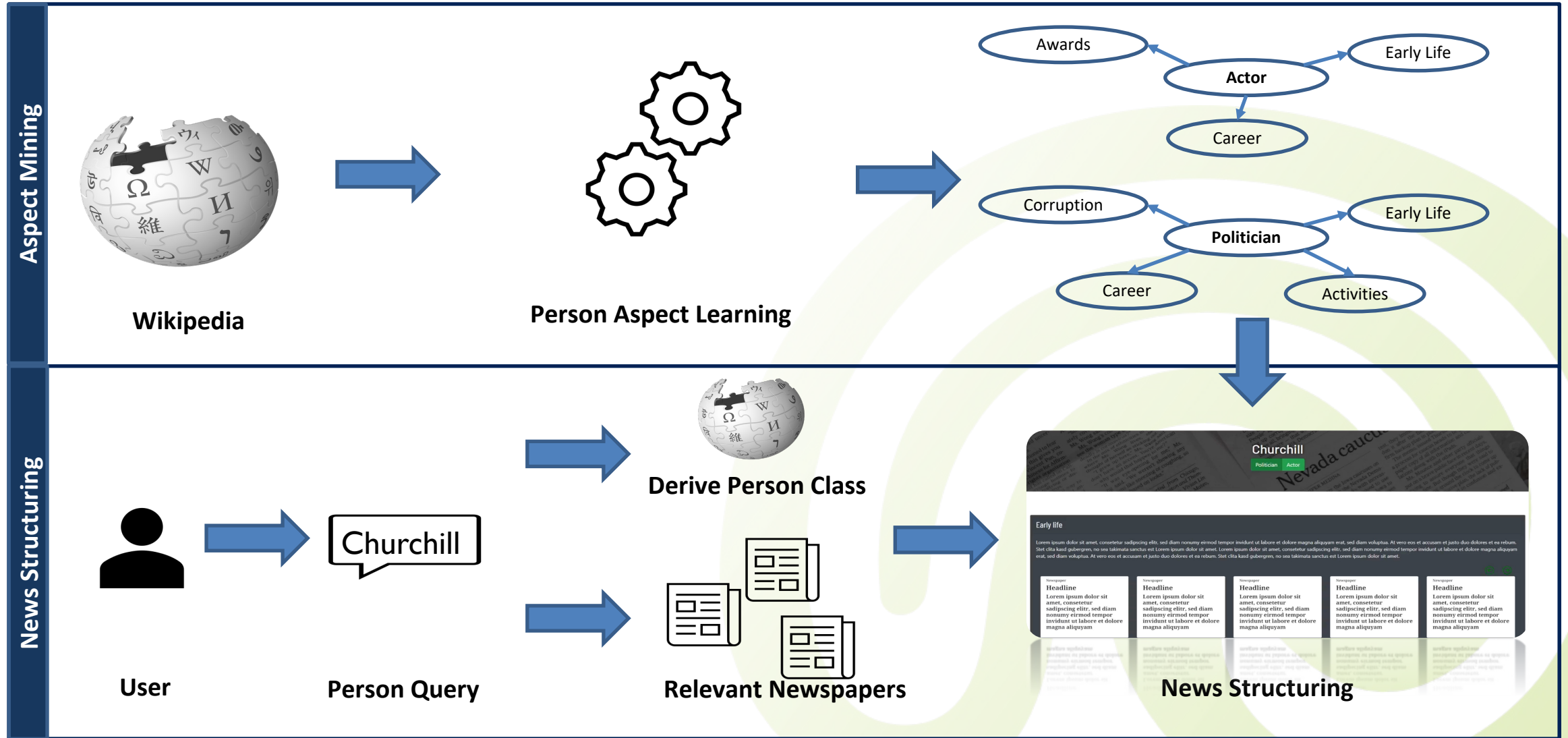
## **De loopbaan van Eisenhower**

Fragment:

[...]President Eisenhower is  
op 14 oktober 1890 in  
Texas geboren. Hij koos  
een militaire loopbaan, die  
in de tweede Wereldoorlog  
zijn bekroning vond in het  
opperbeve[...]



# Systematic Overview





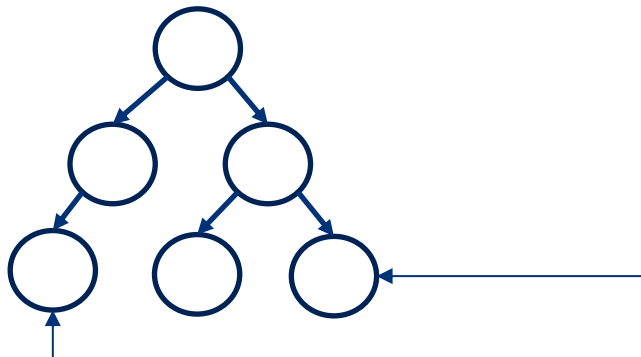
# Deriving Person Classes



## Winston Churchill:

- Brits Minister
- Brits Journalist

## **Person Class Ontology**



**Prune  
&  
Expand**

## Winston Churchill:

- Person
- Minister
- Politician
- Journalist
- Writer





# Learning Person Aspects



## Wikipedia Articles

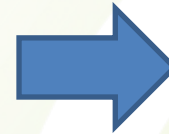
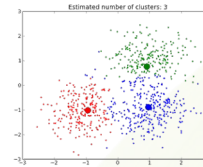
### Frequent Classes:

- Person
- ~~Minister~~
- Politician
- ~~Journalist~~
- ~~Writer~~



### Frequent Aspects:

- Early Life
- Family
- Job

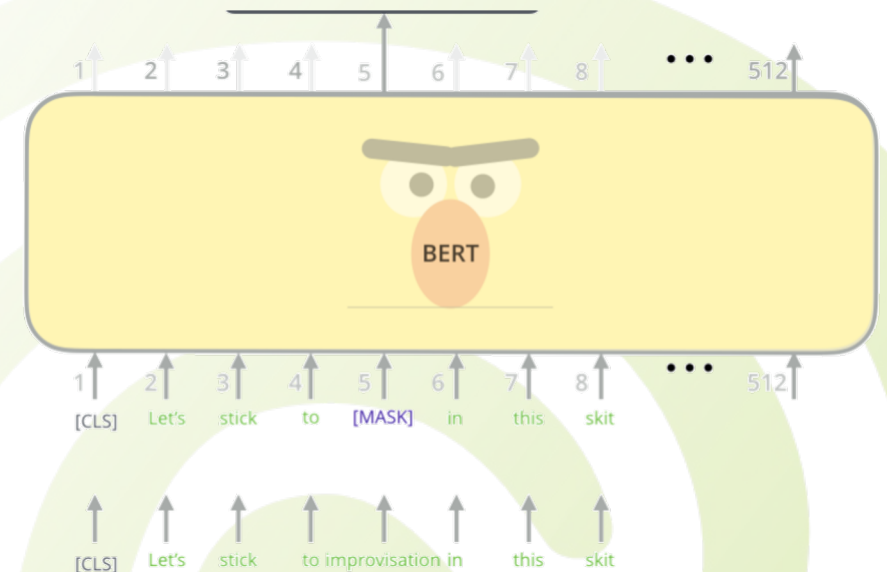


### Frequent Aspects:

- Early Life
- Political Career



## Dutch Language Model



**“Take a text and  
predict an aspect label”**

# Churchill

Politician Actor

## Early life

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Newspaper

### Headline

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam

Newspaper

### Headline

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam

Newspaper

### Headline

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam

Newspaper

### Headline

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam

Newspaper

### Headline

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam

առնաք զիմանալ  
նախընտրուի յարօրե Եւ զօյրե  
նօսնալ Եւնօզ լեւծօր  
զգրիշշուճ Երմւ' շեզ զիսա  
աւե՛ր՝ Կօնշեւեմ  
Լօրեա յնշաա զօյր չի  
Խեղճիւն

առնաք զիմանալ  
նախընտրուի յարօրե Եւ զօյրե  
նօսնալ Եւնօզ լեւծօր  
զգրիշշուճ Երմւ' շեզ զիսա  
աւե՛ր՝ Կօնշեւեմ  
Լօրեա յնշաա զօյր չի  
Խեղճիւն

առնաք զիմանալ  
նախընտրուի յարօրե Եւ զօյրե  
նօսնալ Եւնօզ լեւծօր  
զգրիշշուճ Երմւ' շեզ զիսա  
աւե՛ր՝ Կօնշեւեմ  
Լօրեա յնշաա զօյր չի  
Խեղճիւն

առնաք զիմանալ  
նախընտրուի յարօրե Եւ զօյրե  
նօսնալ Եւնօզ լեւծօր  
զգրիշշուճ Երմւ' շեզ զիսա  
աւե՛ր՝ Կօնշեւեմ  
Լօրեա յնշաա զօյր չի  
Խեղճիւն

առնաք զիմանալ  
նախընտրուի յարօրե Եւ զօյրե  
նօսնալ Եւնօզ լեւծօր  
զգրիշշուճ Երմւ' շեզ զիսա  
աւե՛ր՝ Կօնշեւեմ  
Լօրեա յնշաա զօյր չի  
Խեղճիւն



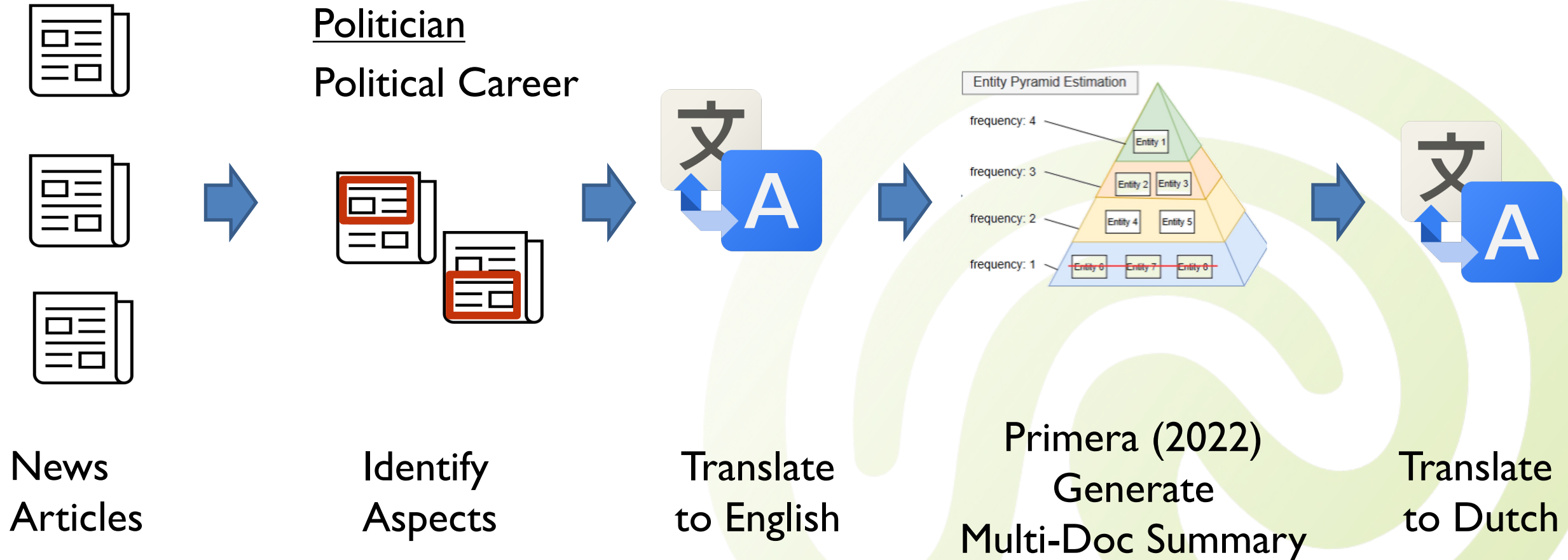
# Summarizing Content

- Apply some fine-tuned multi-document summary model...
  - Unfortunately, we did not find any **fine-tuned** multi-document summary model for **Dutch**





# Multi-Document Summarization







# Evaluation (1/2)

1. **Clustering** of Wikipedia article titles to derive aspects
  - Worked fine
2. **Aspect text classification**
  - Worked fine but diversification required
3. **Translation** Dutch-to-English
  - Minor mistakes but no factual errors



## Evaluation (2/2)

### 4. Multi-Document **Summarization**

- Major issues with **hallucinations**
- “The New York times reports, click here for more”

### 5. **User Demo + Interviews**

- Intuitive + carries much information
- Minor improvements would be beneficial
- “Summaries containing incorrect facts are problematic for a library”



# Conclusion

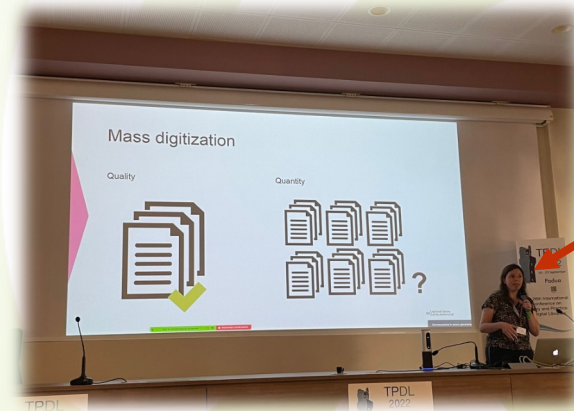
- Go to conferences – it's worth it



**You are here**



**Hermann**

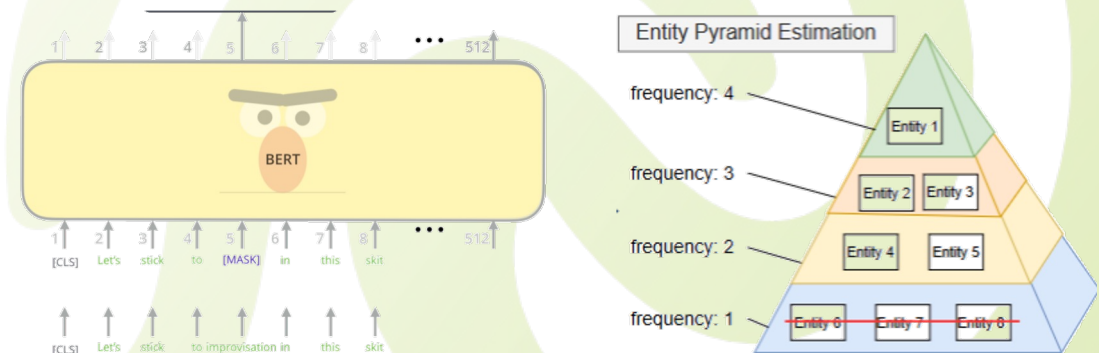
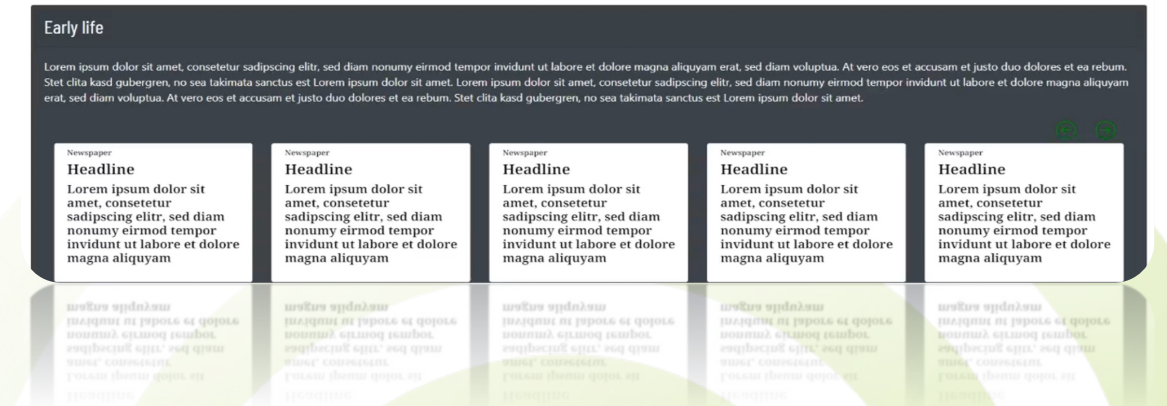
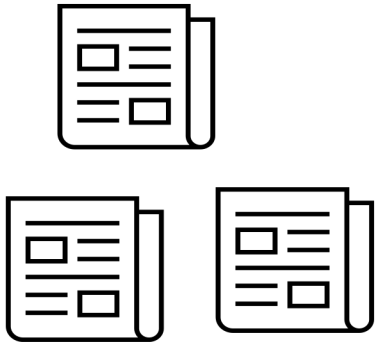


**Mirjam**



# Conclusion

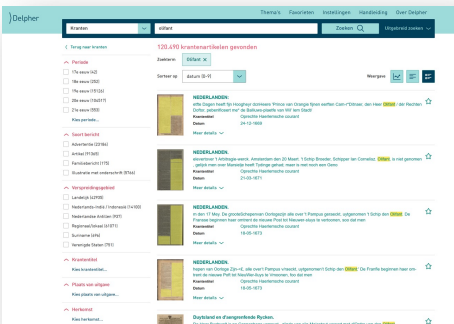
- Our aspect-driven access path effectively allows to explore the KB's content
  - How? By **reusing** existing data and applying modern NLP methods







# Thank You!



KB } national library  
of the netherlands

If you have any questions, contact us:



[kroll@ifis.cs.tu-bs.de](mailto:kroll@ifis.cs.tu-bs.de)



[@HermannKroll](https://twitter.com/HermannKroll)



[@hkroll@fosstodon.org](mailto:@hkroll@fosstodon.org)



[@CuperMirjam](https://twitter.com/CuperMirjam)



[mirjam.cuper@kb.nl](mailto:mirjam.cuper@kb.nl)



<https://github.com/HermannKroll/AspectDrivenNewsStructuring>