



Fluosoft: een AI experiment

- Alain Descamps
- Brocade – Library & Archive Software
- Anet – University of Antwerp Library

anet.be



Fluosoft:

- - De aanleiding
- - Opzet
- - Resultaten
- - Ervaringen
- - Interne werking

Fluosoft: aanleiding

- Springer books :
 - - import ca. 20.000 beschrijvingen
 - - onvolledige data voor UDC bepaling
- CRIS : (Academische Bibliografie)
 - - VODS codes toe te kennen (retro actief ?)
 - (Vlaamse Onderzoeksdisciplines)

→ **Manueel ??**

A view on 'the cataloger'

Informatie (boek, elektronisch)

→ structureren: titel, uitgever, collatie, plaatskenmerken, onderwerp(en), ..

→ coderen :

- taalkode
- titeltype
- onderwerpscodes (UDC, AAT,..)
- personen
- plaatsen
- ...

- **Oorlog en Vrede → UDC 882 (Russische letterkunde)**

A 2nd view on 'the cataloger'

- Thesaurus:

- Personen, Plaatsen, UDC, Trefwoorden,...

- 'Mini' thesaurus:

- Taalcodes

- Titeltypes

- Annotatietypes

- ..



AI try-out : Copilot vraag 1

Voorbeeld: titel:

- *“Language in relation to a unified theory of the structure of human behavior”*

Copilot → **UDC codes** ?

AI try-out : Copilot antwoord 1

Voorbeeld: titel:

- *“Language in relation to a unified theory of the structure of human behavior”*

Copilot: **81: Linguïstiek en talen**

81.1: Algemene linguïstiek

81.23: Taalgedrag en taalgebruik

81.27: Psycholinguïstiek

81.28: Sociolinguïstiek

159.9: Psychologie

159.92: Gedragspsychologie'

(Catalogus Anet : **800.13 Sociolinguïstiek**
15 Psychologie)

AI try-out : Copilot vraag 2

Voorbeeld: titel:

- *“Language in relation to a unified theory of the structure of human behavior”*

Copilot → VODS codes ?

= Vlaamse Onderzoeksdiscipline Standaard

Copilot: conclusies

- Resultaten niet steeds ter zake
 - 'Gesprek' nodig ==> tijdsintensief
 - Bruikbaar na aanlooptijd
 - Interpretatie nodig !
 - Integratie Brocade ?
 - Energieverslindend
 - Black box
-
- (So tell me what you want, what you really, really want)

Fluosoft: aan de slag !

Experiment 'Fluosoft':

1 VTE + Brocade

(Copilot:

1,2 miljoen VTE's)

• = Fluostift + Software

Fluosoft: opzet

1. Leerproces
2. Suggestie algoritme:
~→ codes
3. Black box revelation:
- Visualisatie

- Geen 'natural' language omgeving

Fluosoft: Leerproces (1)

- Link benamingen met codes:
 - AAT (Art & Architecture Thesaurus)
 - UDC
 - Geografische codes
- Hoe ?

• = Voorbeeld “*Russische letterkunde*” → UDC 882

Fluosoft: Leerproces (1b)

- Link woorden uit benamingen van codes met de codes:
 - AAT (Art & Architecture Thesaurus)
 - UDC
 - Geografische codes

• = woorden “*Russische*” “*letterkunde*” → UDC 882

Fluosoft: Leerproces (3)

- Link Titels in catalogus met
 - toegekende **taalcodes**

• = Voorbeeld: “*Oorlog*” “*en*” “*Vrede*” → dut (Nederlands)

Fluosoft: Suggestie algoritme (2)

- Idee:

1. Zoek met belangrijkste woord
2. Zoek met 2 belangrijkste woorden
3. ...
4. Stop als volgende stap niets oplevert

• → **Suggestie <> Zoekactie !**

Fluosoft: Black box revelation

- Visualisatie:
 - 'Fluoresceer' belangrijke woorden
- Uitgebreid technisch rapport

Fluosoft: Black box revelation

- Visualisatie:

'Fluoresceer' belangrijke woorden: voorbeeld:

Case study, based on a critical discourse analysis of news articles published in Dutch-speaking Belgian news media

Fluosoft: Black box revelation

- Uitgebreid technisch rapport

- **threshold is 83156=15% of 554371**
Removing word a (139667>83156)
Removing word of (97094>83156)
looking for codes in restricted selection on level 2
subgroup 1 has total of 554371 hits
text:analysis has 60 hits
text:articles has 58 hits
text:based has 58 hits
...

Leerproces: woordenslag (1)

Voorbeeld: UDC 372.87: 'Didactiek van de kunst'

'Didactiek' → code '372.87'

'van' → code '372.87'

'de' → code '372.87'

'kunst' → code '372.87'

- **(Herhaal 962 maal)**

Leerproces: woordenslag (2)

Resultaat:

'Didactiek' → 39 hits:

-> '372.83' (Didactiek van de maatschappelijke vorming)

-> ...

-> '372.87' (Didactiek van de kunst)

'van' → 118 hits

'de' → 155 hits

'kunst' → 3 hits

- **(Hit the road, Jack)**

Leerproces: woordenslag (3)

- Rangschikking: minder hits = specifiekere = belangrijker:
- 1. kunst (3)
- 2. didactiek (39)
- 3. van (118)
- 4. de (155)



Leerproces: woordenslag (4)

- Negeer 'stopwoorden':

- 1. kunst (3)

- 2. didactiek (39)

- ~~▪ 3. van (118)~~

- ~~▪ 4. de (155)~~

- (won't you come back no more.)

Leerproces: woordenslag (2)

Associeer titels/abstracts

met toegekende UDC codes

Voorbeeld: c:lvd:112145: "***Kunst und Unterricht***"

– *kunst* → code '372.87' (Didactiek van de kunst)

– *und* → code '372.87' (Didactiek van de kunst)

– *unterricht* → code '372.87' (Didactiek van de kunst)

- **(Herhaal >3.000.000 maal)**

Leerproces: kost

UDC's van 3.325.261 beschrijvingen:

- Collect
 - Maak suggesties
 - Vergelijk suggesties met realiteit
 - in 7h 52 min
 - 117 / seconde
-
- (Analoog voor AAT, taalcodes,..)

Leerproces: kost (2)

UDC's van 3.325.261 beschrijvingen:

- Sneller dan Copilot, ChatGPT,..
- Zuiniger
- Beter ?

- (You can't say we never tried..)

Ervaringen ..

Suggestie taalcode (1)

Tekst

Case study, based(?) on a critical discourse analysis of news articles published(?) in Dutch-speaking(?) Belgian(?) news media

Suggesties

eng = Engels (76 %)

Suggestie Anet-UDC

Tekst

Case study, based on a critical discourse analysis of news articles published in Dutch-speaking Belgian news media

Suggesties

00.800 = Zonder onderwerpscode: linguïstiek (100 %)

301.54 = Sociologie van het georganiseerde cultuurleven

658.81 = Consumentengedrag [a::658.81] (100 %)

800.05 = Grammatica [a::800.05] (100 %)

800.13 = Sociolinguïstiek [a::800.13] (100 %)

Suggestie AAT : materialen

Tekst

Een levensgroot(?) **reliëf** in verchromd(?) **staal**

Suggesties

300010357 = Materialen (67 %)

300264091 = Facet Materialen (67 %)

→ Nivellering nodig !

Suggestie AAT : materialen (met 'nivellering')

Tekst

Een levensgroot(?) **reliëf** in **verchromd(?)** **staal**

Suggesties

300010921 = austenitisch roestvrij staal (67 %)

300010922 = stainless steel, martensitic (67 %)

Suggestie Geografische code

Tekst

Brugge en Oostende

Suggesties

a::91.493.032 = West-Vlaanderen (100 %)

Suggestie taalcode (2)

Tekst

Berlin

Suggesties

eng = Engels (100 %)

ger = Duits (100 %)

Suggestie UDC code – abstract

Tekst

Although they play a fundamental role in nearly all branches of mathematics, inequalities are usually obtained by ad hoc methods rather than as consequences of some underlying "theory of inequalities." For certain kinds of inequalities, the notion of majorization leads to such a theory that is sometimes extremely useful and powerful for deriving inequalities. Moreover, the derivation of an inequality by methods of majorization is often very helpful both for providing a deeper understanding and for suggesting natural generalizations.

Suggesties

530.02 = Mathematische fysica (100 %)

621.38 = Elektronica (100 %)

Fouten in catalogus (1)

Tekst

Jugenderinnerungen(?) **eines** **alten** **Mannes**

Suggesties

ger = Duits (75 %)

Catalogus (c:1vd:800046)

fre |

Foutieve suggesties (1)

Tekst

Jurnal(?) Ekonomi(?) Manajemen(?) Sistem(?)
Informasi(?) (Tangerang.Online)

Suggesties

ger = Duits (17 %)

Catalogus (c:lyd:16017430)

rum

Bumps in the road

- Slechte data in leerproces → onbruikbaar (begin opnieuw !)

Voorbeelden:

- c:lvd:6601246 → “wel” = Engels ?
- c:lvd:14653768 “*Been there, done that*” → Nederlands ?
- Automatisch leerproces versus kwaliteitscontrole bij leren
- Soms weinig trainingsmateriaal (vb: VODS)

Bruikbaarheid ?

- Suggesties: ja
- Auto-encoding: nee
- Bijsturing nodig

• (should I stay or should I go?)

Mogelijke toepassingen in workflow:

- **Suggestieknop** onderwerpscodes in 'bibrec'.
- **Controle diversiteit van benamingen**
 - Vb: VODS codes
 - *Optimisation [01010903]* versus
 - *Structural optimisation [02010102]*
 - Vb : UDC codes
 - *Geschiedenis van het klassieke Griekenland UDC 938*
 - *Geschiedenis van Griekenland UDC 949.5*
- **Automatische herkenning stopwoorden**



Mogelijke toepassingen in workflow(2):

- **Publiek toegankelijk suggestiescherm voor:**
 - AAT codes : materialen, stijlen,..
 - Zelflerende kennisomgeving ?
 - UDC codes
 - VODS
 - Eigen trefwoordensysteem ?

- **(Riders of the (brain)storm..)**

Beschikbaar release 6.40:

- <https://anet.be/menu/fluosoft>
- Of zoek naar 'Fluosoft' in menu.
- Kies een kennisomgeving
- Geef een titel/abstract in
- Druk op 'Registreer'



Discussie

Vragen – Bemerkingen ?

- Alain Descamps
- alain.descamps@uantwerpen.be