

Toepassingen van intelligente informatieverwerking in E-Discovery

Author(s)

Henseler, Hans

Publication date

2011

Document Version

Final published version

[Link to publication](#)

Citation for published version (APA):

Henseler, H. (2011). *Toepassingen van intelligente informatieverwerking in E-Discovery*. Paper presented at Symposium E-Discovery in Nederland, Amsterdam.

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please contact the library: <https://www.amsterdamuas.com/library/contact>, or send a letter to: University Library (Library of the University of Amsterdam and Amsterdam University of Applied Sciences), Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Toepassingen van intelligente informatieverwerking in E-Discovery

Hans Henseler

*Symposium E-Discovery in Nederland,
Amsterdam 11 mei 2011*

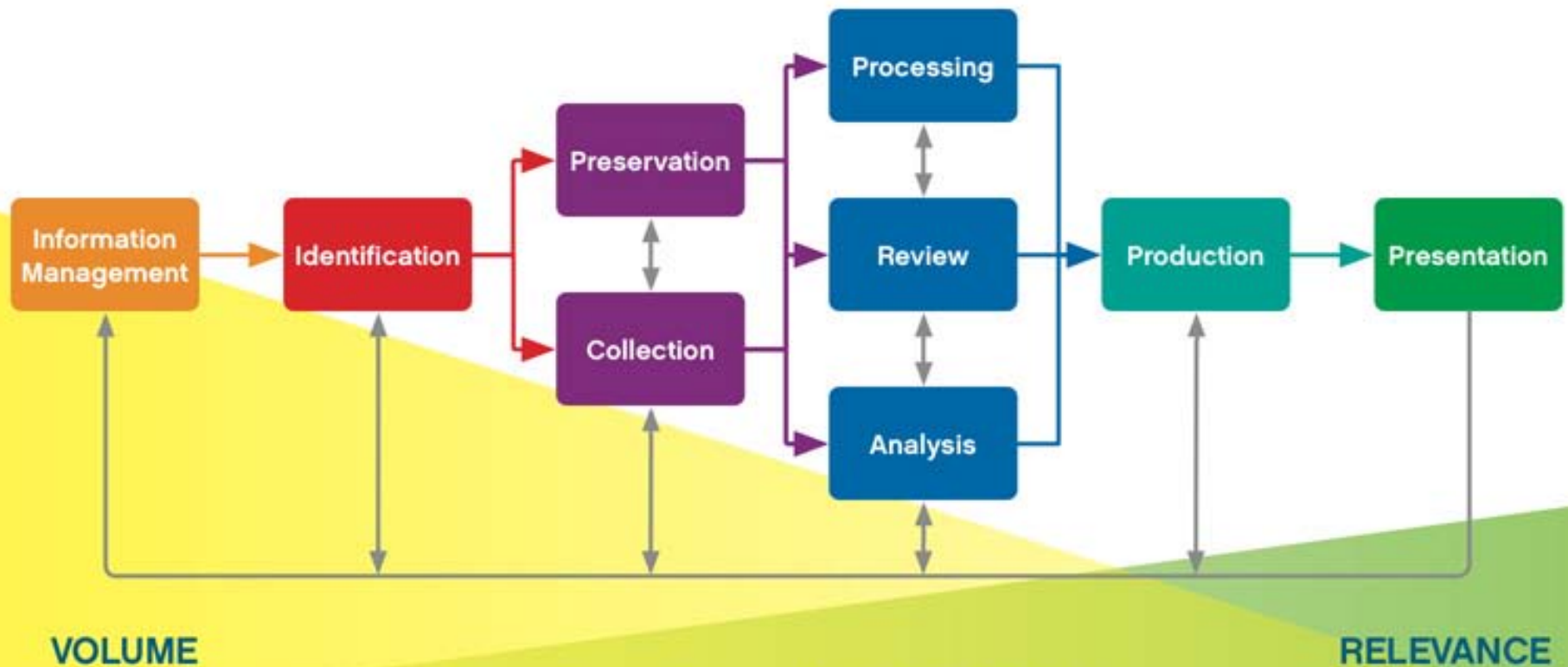
Email: j.henseler@hva.nl Twitter: @ediscoverynl



Inhoud

- **E-Discovery Reference Model**
- **Recall en Precisie in de praktijk**
- **Transparant zoeken**
- **Relevance Ranking**
- **Visualisatie**
- **Email threading**
- **Automatische review**
- **Zoeken met meta data**
- **Entity Extractie**
- **Topic maps**
- **Onderzoek aan de HvA en UvA**

Electronic Discovery Reference Model



Electronic Discovery Reference Model / © 2009 / v2.0 / edrm.net

Recall en precisie in de praktijk

- **Verbeteren van Recall:**
 - Wildcards, Stemming and Fuzzy Search
 - OR operator
- **Verbeteren van Precisie:**
 - AND operator
 - Proximity Search
 - Preview query results on
- **Geavanceerde technieken:**
 - Transparante search

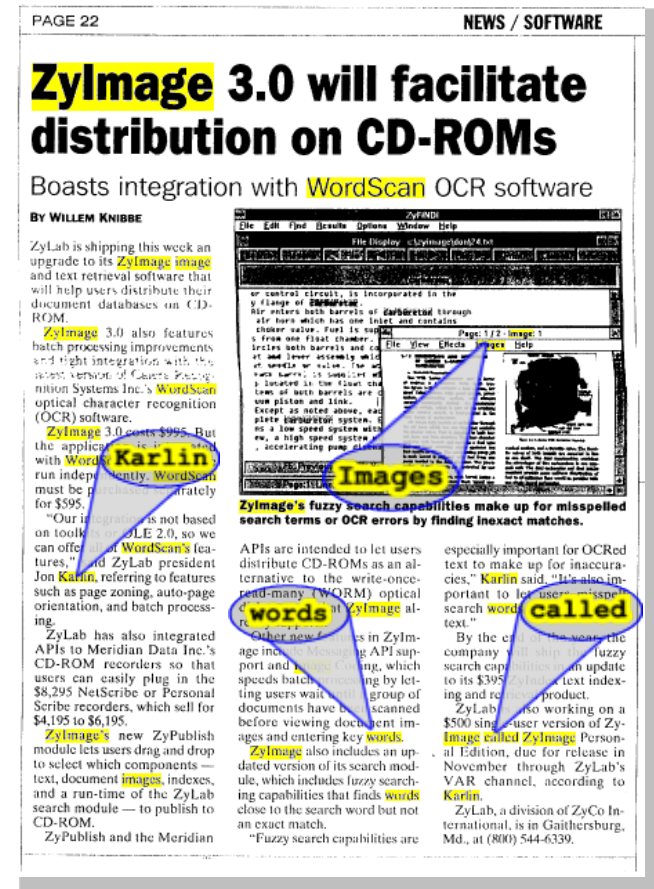
Fuzzy Zoeken — nadruk op hogere recall ten koste van de precisie

Voorbeeld

Zoeken naar:

- “Syimage”
- “Wordscan”
- “Carlin”
- “Calera”

Zelfs in tekst van lage kwaliteit (bijv tgv OCR) zal fuzzy search matches en spelfouten kunnen corrigeren.



Transparant zoeken

Select Variations

Select variations for: fals* Show: All

☐	Variation	Matching Emails	Matching Unique Files
☐	falsely	12	6
☐	falseness	74	5
☐	falsestart	274	4
☐	falseetto	74	5
☐	falsie	0	5
☒	falsification	1	6
☒	falsifications	15	4
☒	falsified	68	8
☒	falsify	27	0
☒	falsifying	29	8
☐	falsity	148	5
☐	falsom	43	2

Displaying 8 - 19 of 21

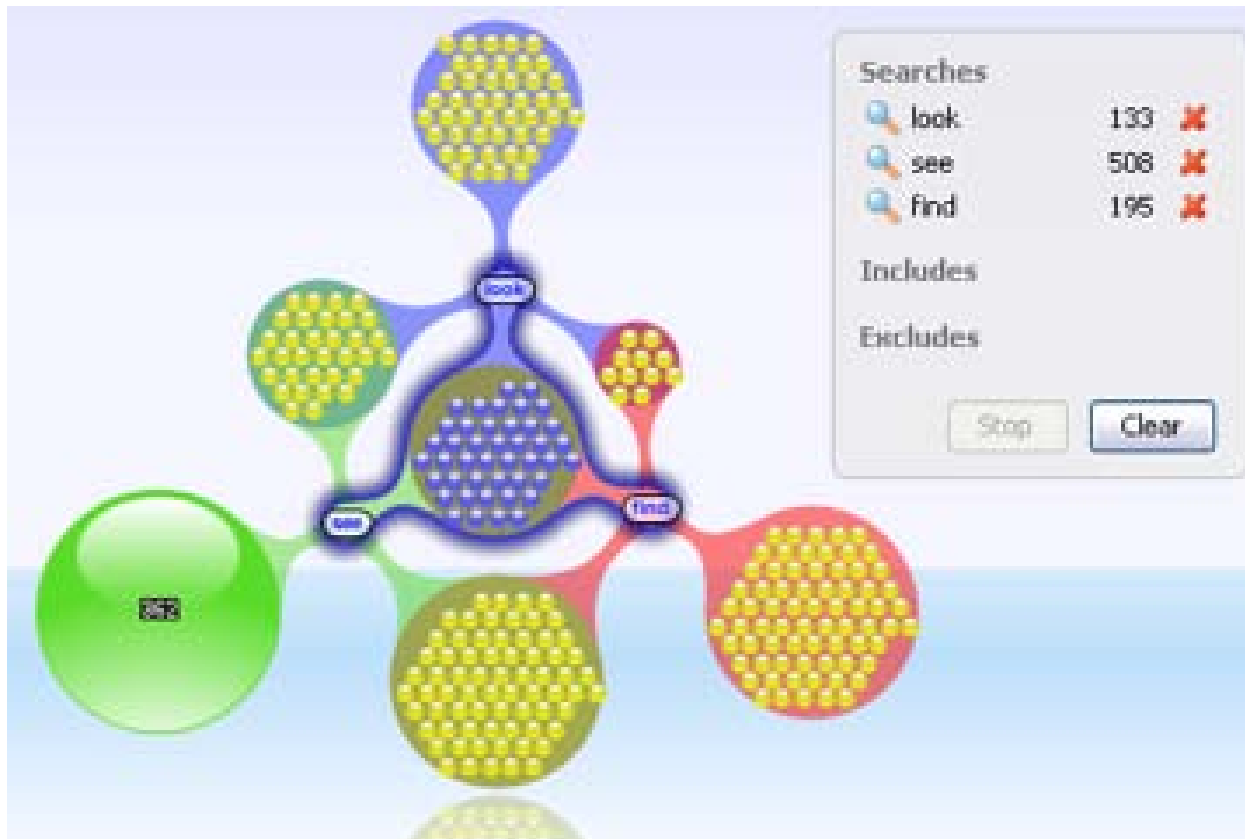
OK Copy Shown Variations Cancel

Relevance ranking

- Google pageRank :



Visualisatie



E-mail threading

Discussion Analysis: Wall Street Earnings Numbers

Summary Thread Terms Conversations Senders Attachments Analyzed across case

Chronological Thread of the Discussion: Wall Street Earnings Numbers
Note: Derived messages may be shown below for discussion thread clarity, but cannot be tagged or exported.

Tag Export Print Cache (0 of 15 selected)

	Sender	Reci
☐	Phil Smith	B
☐	Bernie Royan	
☐	Mike Tamas	
☐	Bernie Royan	
☐	David Moore	
☐	Jenny Richards	
☐	Phil Smith	
☐	Sally Burton	
☐	Robert Berg	
☐	Mike Tamas	
☐	Mike Tamas	
☐	Mike Simpson	

Extend view |<< < Prev | 1 | Next > >> X

Doc ID: 0.7.17.2634
Custodian: Tamas, Mike
Language: English (100%)

From: Phil Smith <philsmith@tamascorp.com>
To: Bernie Royan <bernieroyan@tamascorp.com>
Cc: Jenny Richards <jennyrichards@tamascorp.com>; Robert Berg <robertberg@tamascorp.com>; Matt Stevens <mattstevens@tamascorp.com>
Sent: Thu Sep 28 2006 09:39:00 EDT
Subject: Wall Street Earnings Numbers
Attachments: (1)

Q3 Rev.xls (22 KB) (18)

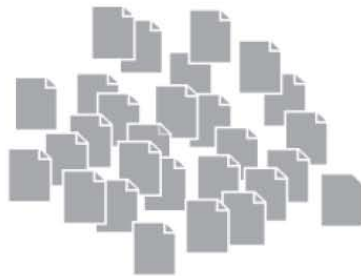
Bernie,

I am confident we can make the earnings number. Revenue expectations were inflated last quarter, but we should be ok. Take a look at the East Coast numbers and tell me if they match up with what the regional managers can defend out in their respective territories.

Based on this data (see attached file), we should be able to meet the \$145M forecast we promised to the street. Also keep this in mind... if we miss, the board will most certainly scrutinize our logic that we articulated at the last board meeting. Talk to you tomorrow.

|<< < Prev | 1 | Next > >>

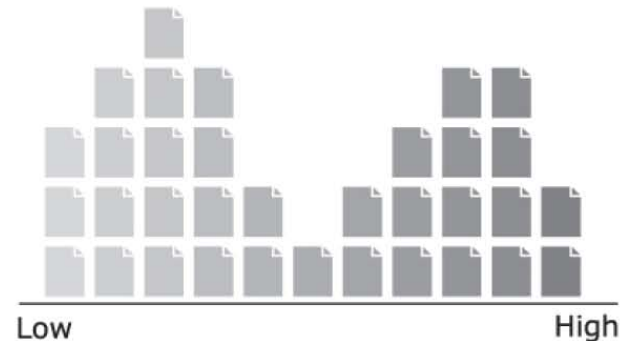
“Automatische” review



Unstructured data collection



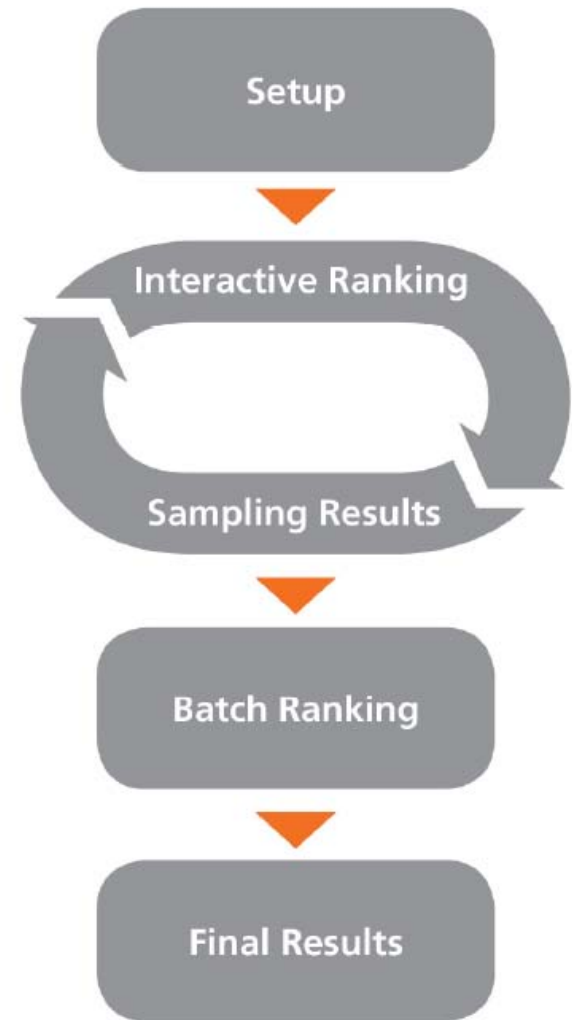
Expert reviews sample documents for relevance



Software calculates relevance scores for documents

Predictive coding

- **Voordelen:**
 - Transparantie
 - Reproduceerbaarheid
 - Geheimhouding
 - Snelheid



Zoeken met meta data

- <http://www.autoscout24.nl/>

Tweedehands auto's  

Merk	Model
Alle	Alle
Verkoopprijs (€) tot	Bouwjaar van
Alle	Alle
Kilometerstand tot	Brandstof
Alle	Alle
Postcode	Omgeving
	100 km

☐ Occasion garantie

[Uitgebreid zoeken](#)

1.896.016 auto's

 **Zoeken**

- ☒ **Merk**
 - [Volkswagen](#) (14233)
 - [Renault](#) (13942)
 - [Opel](#) (12190)
 - [Ford](#) (11502)
 - [Peugeot](#) (10645)
 - [Meer](#)
- ☒ **Brandstof**
 - [Benzine](#) (121901)
 - [Diesel](#) (37154)
 - [Meer](#)
- ☒ **Carrosserievorm**
 - [4/5-Deurs](#) (62999)
 - [2/3-Deurs](#) (28373)
 - [Stationwagen](#) (23073)
 - [Meer](#)
- ☒ **Vermogen kW (PK)**
 - [44-81 \(60-110\)](#) (75915)
 - [81-110 \(110-150\)](#) (48331)
 - [110-257 \(150-349\)](#) (29898)
 - [Meer](#)
- ☒ **Opties**
 - [Stuurbekrachtiging](#) (139806)
 - [Centrale vergrendeling](#) (132896)
 - [Elektrische ramen](#) (127863)
 - [ABS](#) (127533)
 - [Airbag](#) (124828)
 - [Meer](#)
- ☒ **Transmissie**
 - [Handgeschakeld](#) (129190)

Entity Extractie

Entity Extractor: vastgoed

File Process

HAARLEM (AFN) - Eric Frische zat in het beklagdenbankje naast de zwijgzamere medeverdachte **Will Frencken**, oud-directeur van de vastgoedbeheerder van **Philips Pensioenfonds**. Het **Openbaar Ministerie (OM)** verdenkt de voormalig accountant in het grote vastgoedfraudeproces **Klimop** van onder meer het witwassen van meer dan **10 miljoen euro**. **Will Frencken** wordt onder andere verdacht van het verduisteren van geld van **Philips Pensioenfonds**.

Eric Frische zei in **2001** over het pakket onroerend goed te zijn benaderd door **Will Frencken**. Deze vertelde hem dat **Hilders** belangstelling had en dat de vastgoedman er wel wat „tipgeld” voor overhad. **Hilders** mocht niet weten dat **Will Frencken** achter de tip zat, zei Eric Frische.

Daarom sloot hij met zijn bv **Universum** een overeenkomst met **Hilders' bv Ceylonstaete**. Daarin stond dat **Universum** recht had op de helft van de verwachte winst op het project van **24 miljoen gulden**. Voor dat bedrag had **Universum** weinig gedaan, erkende Eric Frische. „Het was eigenlijk tipgeld”, zo verklaarde hij. Het vastgoed werd in **2002** voor in totaal **450 miljoen gulden** aan **Ceylonstaete** verkocht. Volgens het OM voor een te laag bedrag.

Het was voor buitenstaanders niet zichtbaar dat **Will Frencken** eigenlijk de grootste aandeelhouder was van **Universum**, en daarmee het grootste deel van het „tipgeld” kreeg. Accountant **Eric Frische** had een speciale constructie bedacht op verzoek van **Will Frencken**. **Frencken** lag in een echtscheiding en vreesde dat zijn ex beslag zou laten leggen op zijn aandelen, zo zei hij. **Will Frencken** beaamde dat.

Eerder verklaarde Eric Frische tegen de opsporingsdienst **FIOD** dat hij „eigenlijk ontzettend dom” is geweest ertussen te gaat zitten, omdat **Will Frencken** niet in beeld wilde. Eric Frische nam **maandag** afstand van zijn eerste vier **FIOD**-verhoren.

Het proces gaat **donderdag** verder met de inhoudelijke behandeling van de zaak van hoofdverdachte **Jan van Vlijmen**. **Harry Hilders** ontsprong de dans. De vastgoedman, die werd verdacht van omkoping, schikte vorig jaar voor **40 miljoen euro** met het OM en **Philips Pensioenfonds**. Ook kreeg hij een taakstraf van **120 uur**.

Confidence Entity

Name

0,42	Frencken. Frencken
0,77	Jan van Vlijmen

NameAlgorithmic

0,71	AFN
0,23	Will Frencken
0,60	Philips Pensioenfonds
0,60	Openbaar Ministerie
0,45	Klimop
0,45	Universum
0,45	Ceylonstaete
0,47	Accountant Eric Frische
0,71	FIOD
0,40	Harry Hilders

TimeReference

0,90	maandag
0,90	donderdag

GeographicReference

0,90	HAARLEM
------	---------

Quotation

1,00	tipgeld
------	---------

Number

1,00	10
1,00	2001
1,00	24
1,00	2002
1,00	450
1,00	eerste
1,00	vier
1,00	40
1,00	120

NEWSMAP IS ©2004-2010 MARCOS WESKAMP | [CONTACT](#) | [BLOG](#) | [ABOUT](#)

powered by hosting by



Hogeschool van Amsterdam
Media, Creatie en Informatie

Onderzoek HvA en UvA

- **Voorstel ingediend bij NWO voor semantisch zoeken in E-Discovery (2 AIO's)**
- **Promotievoorstel in voorbereiding bij de HvA/UvA: Identificatie en extractie van entiteiten en relaties tussen entiteiten in het kader van:**
 - Personen en feiten die te maken hebben met omkoping
 - Personen en feiten die te maken hebben met het overtreden van de mededingingswet
 - Identificatie van privé- en geheimhoudersinformatie

IBM Watson: Vraag en Antwoord



