

---

# Rol van de kennisinfrastructuur voor het bedrijfsleven

---

Piek Vossen, Irion Technologies

Stevin

2-maart-2005, Antwerpen

---

# Overzicht

- Wat voor bedrijf is Irion Technologies?
- David & de Goliath(s)
- Innovatie en taaltechnologie tegen het grootkapitaal.
- Samenwerkingsprojecten met universiteiten:
  - Meaning
  - Pidgin

---

# Irion Technologies

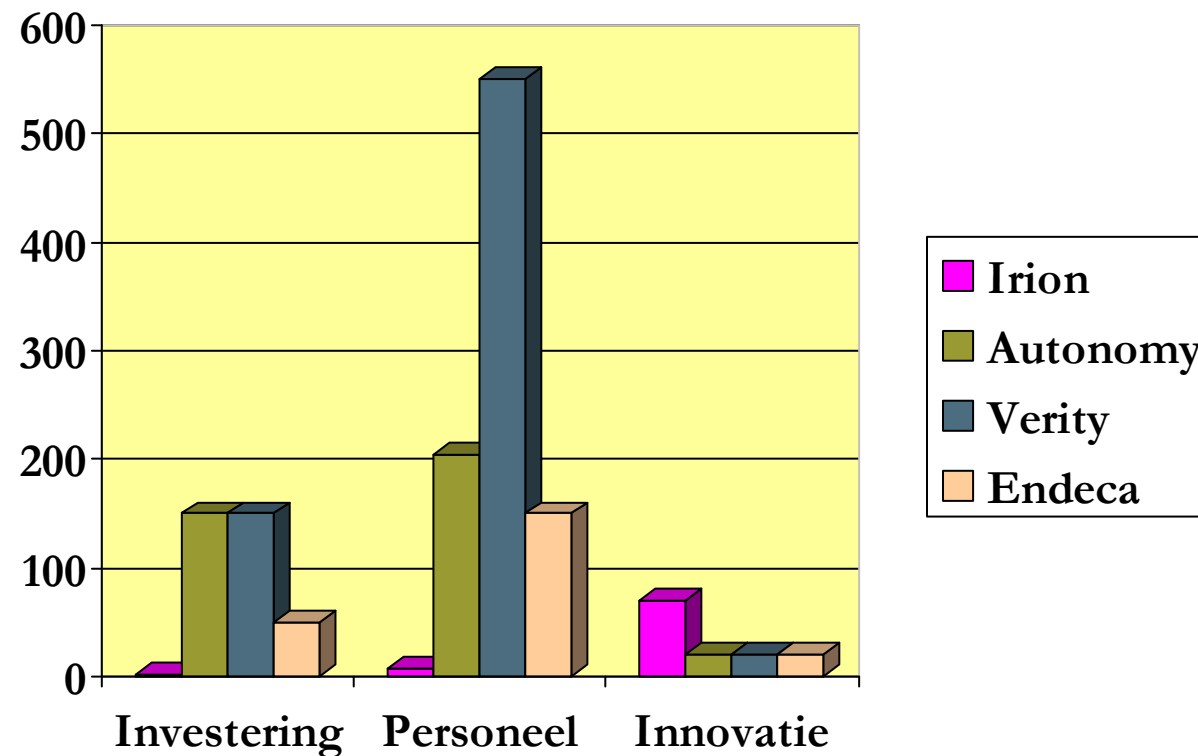
- Klein innovatief bedrijf
- Sterke technologische basis:
  - Software en technologie van TNO
  - Lingware van Van Dale
- Hoofdproducten:
  - Cross-lingual retrieval (6 talen direct, meer talen op verzoek)
  - Document classificatie
  - Dialoog systemen

---

# David & de Goliath(s)

- Grootste concurrenten in Nederland
  - Autonomy, Verity, Endeca
  - Amerikaanse bedrijven
  - Veel kapitaal en al lang actief op de markt
- Ontevreden klanten:
  - Statistische technieken niet goed genoeg;
  - Nederlands morfologisch complexer dan Engels;
- Speciale eisen:
  - Multilingualiteit & cross-lingualiteit;
  - Meer recall en meer precision;
  - Complexe inhoudelijke analyse en kennisautomatisering;

# Innovatie en taaltechnologie tegen het groot-kapitaal



---

# Samenwerking voor innovatie en marktontwikkeling

- Ontwikkel cutting-edge technologie:
  - EU project MEANING:
    - Automatische massale acquisitie van lexicale kennis uit tekstcorpora voor 5 talen;
    - Ontwikkeling van WSD technologie op grond van die kennis;
    - Cross-linguale kennisdeling via EuroWordNet database;
  - Senter project Pidgin:
    - Zelflerend cross-linguaal chatsysteem;
- Rol van Irion:
  - Implementatie in realistische applicaties en omgevingen;
  - Oplossen van echte problemen;

---

# Wat heeft het opgeleverd?

- **MEANING:**

- Significante verbetering van tekstclassificatie door middel van taaltechnologie en woord-desambiguatie;
- Nieuwe methode van conceptueel zoeken;
  - Significante verbetering van mono- en cross-linguale retrieval;
  - Verdubbeling van de productiviteit van eindgebruiker;

- **Pidgin:**

- Spin-off: nieuw product waarin informatiezoekers door middel van een zelflerende dialoog naar de juiste plaats worden geleid;
- Werkt op ongestructureerde data en heeft geen wereldmodel nodig;

---

## Text-classification on Reuters news

	Recall	Precision	Coverage
No Wordnet	67.7	70.4	83.2
Wordnet	75.6	65.9	99.5
Wordnet + WSD	<b>79.2</b>	<b>71.5</b>	<b>100</b>



# Retrieval of pictures from EFE database

Multi word queries	Spanish original 105 Q		Spanish paraphrase 94 Q		English 105 Q		Catalan 105 Q		Basque 104 Q		Italian 105 Q	
	Count	Score	Count	Score	Count	Score	Count	Score	Count	Score	Count	Score
<b>NO</b>	99	0.94	14	0.15	2	0.02	31	0.3	1	0.01	3	0.03
<b>p1</b>	60	0.57	9	0.1	0	0	21	0.2	1	0.01	2	0.02
<b>FULL</b>	96	0.91	71	0.76	39	0.37	70	0.67	50	0.48	39	0.37
<b>p1</b>	55	0.52	38	0.4	16	0.15	44	0.42	27	0.26	19	0.18
<b>MEANING</b>	97	0.92	61	0.65	39	0.37	68	0.65	46	0.44	32	0.3
<b>p1</b>	60	0.57	39	0.41	21	0.2	48	0.46	27	0.26	20	0.19

# End user task to find pictures

	NO	FULL	MEANING
SEARCH	110	64	<b>56</b>
HIGHLIGHT	105	<b>55</b>	60
DISAP.	57	28	<b>27</b>
CONFIRMED	20	19	<b>24</b>
UNDEC.	3	6	<b>1</b>
TOTAL	295	172	<b>168</b>

---

Dank u voor uw aandacht!

---