

# Martin Henk van den Berg

650-814-0645 — mhvdberg@gmail.com — <https://www.linkedin.com/in/vdberg/>

Data scientist with extensive experience in the area of document analysis for the purposes of summarization, knowledge extraction and search. Managed groups developing and deploying natural language components within research and product divisions. Developed both back-end and front-end tools and technology for Natural Language applications. Designed knowledge infrastructure for co-reference resolution in mixed-initiative dialog systems.

Architected a multi-year knowledge extraction project in cooperation with the Natural Language and Knowledge Representation groups at Xerox PARC that eventually spun-out as the startup Powerset.

Co-designed “Data-Oriented Parsing”, one of the first big-data parsing projects, with Remko Scha.

Co-authored 24 granted patents.

## EXPERIENCE

### RESEARCH SCIENTIST

**INTEL CORP., SANTA CLARA, CA — 2012–present**

- Managed a group developing language tools and technologies for Oakley’s Radar Pace: sunglasses with a build in running and cycling coach and the general purpose successor Dialog Express.
- Evaluated a number of external companies for potential for acquisition or investment, leading to two acquisitions (at the request of Intel Capital).
- Member of the Security patent committee.

### SENIOR RESEARCH SDE

**MICROSOFT CORP. — SVC, MOUNTAIN VIEW, CA — 2008–2012**

- Maintained and further developed core Natural Language Tools and Technologies licensed from Xerox PARC: Finite State technology, syntactic parser (XLE) and semantic processing component (“Transfer”). Integrated FST into Bing’s Whole Page Relevance query-classifier and Cortana back-ends.
- Developed summarization technology for use in the commerce portal within Bing.
- Improved front-end query classifiers for Bing.

### SENIOR SCIENTIST

**POWERSET, SAN FRANCISCO, CA — 2006–2008**

- Helped set up the company starting in 2005 as the first regular employee. Wrote the first demos to the investors.
- Redesigned the document search and indexing technology developed at FXPAL into a Wikipedia search engine.
- As part of the semantics team developed rules to analyze the syntactic parser output and produce indexable facts encoding semantic structure of the indexed texts.
- Developed Ruby libraries for index and query side pipelines to invoke syntactic and semantic processing components written in C, C++, TCL/TK and Prolog. Assisted in port of transfer component from Prolog to C++

### (SENIOR) RESEARCH SCIENTIST

**FXPAL, PALO ALTO, CA — 1998–2006**

- Technical lead and system architecture designer and implementer of a number of experimental text-based natural language applications, including a spoken dialog system and a text-summarization system.
- Working closely with Xerox PARC, developed a linguistic structure preserving index system to use as the core of a document search and question-answering system to support more precise archiving and retrieval of business document as part of Sarbanes-Oxley compliance.

POSTDOCTORAL RESEARCH FELLOW  
**IBM, ALMADEN RESEARCH CENTER, SAN JOSE, CA — 1997–1998**

With Soumen Chakrabarti and Byron Dom, developed "Focused Crawling" algorithms that improve on basic web-crawling by looking at the relevance of the content of documents during the web-crawl. The research won the award for Best Paper at the Eighth International World Wide Web Conference in 1999 and was also one of two papers to win the 1999 IBM Computer Science Best Paper Award.

POSTDOCTORAL RESEARCH FELLOW  
**IBM, SANTA TERESA, SAN JOSE, CA — 1996–1997**

Technical leader of a four person design and implementation effort of a natural language interface for database queries (as part of DB2 development).

VISITING SCHOLAR  
**CSLI, STANFORD UNIVERSITY, STANFORD, CA — 1995**

Continued research into the formal aspects of discourse structure and semantics.

RESEARCH ASSISTANT  
**DEPARTMENT OF COMPUTATIONAL LINGUISTICS, UNIVERSITY OF AMSTERDAM, AMSTERDAM, NL — 1989–1996**

- Did research into the formal aspects of discourse structure and semantics leading to my PhD.
- Taught Topics in Computational Linguistics, Introduction to First Order Logic and Advanced C Programming.

## **PROFESSIONAL SERVICES**

MEMBER OF THE AD HOC ANSI COMMITTEE ON ONTOLOGY STANDARDS — 1996–1998

MEMBER OF THE GOVERNING BOARD OF THE DEPARTMENT OF COMPUTATIONAL LINGUISTICS  
**UNIVERSITY OF AMSTERDAM — 1991–1996**

## **Education**

PHD — COMPUTATIONAL LINGUISTICS  
**UNIVERSITY OF AMSTERDAM, AMSTERDAM, THE NETHERLANDS — 1996**

Thesis: *Some Aspects of the Internal Structure of Discourse. The Dynamics of Nominal Anaphora.*  
Supervisors: Professors Remko Scha and Johan van Benthem.

B.S./M.S ("DOCTORAAL" DEGREE) — THEORETICAL PHYSICS  
**UNIVERSITY OF AMSTERDAM, AMSTERDAM, THE NETHERLANDS — 1989**

Thesis: *Invariants of Transformations. Comparing Relativistic and Non-relativistic Space-Time Structures.*

# PATENTS

## MICROSOFT (INVENTIONS DEVELOPED AT POWERSET)

Identification of semantic relationships within reported speech. USP 9,449,081, 09-20-2016

Identification of semantic relationships within reported speech. USP 8,868,562, 10-21-2014

Coreference resolution in an ambiguity-sensitive natural language processing system. USP 8,712,758, 04-29-2014

Fact-based indexing for natural language search. USP 8,639,708, 01-28-2014

Calculating valence of expressions within documents for searching a document. USP 8,346,756, 01-01-2013

Indexing role hierarchies for words in a search index. USP 8,229,730, 07-24-2012

Semi-automatic example-based induction of semantic translation rules to support natural language search. USP 8,041,697, 10-18-2011

## FXPAL

Cascading cluster collages: visualization of image search results on small displays. USP 7,904,455, 03-8-2011

Systems and methods for determining relevant information based on document structure. USP 7,739,279, 06-15-2010

Systems and methods for brokering services. USP 7,689,645, 03-30-2010

Systems and methods for hybrid text summarization. USP 7,610,190, 10-27-2009

Systems and methods for collaborative note-taking. USP 7,542,971, 06-2-2009

Systems and methods for determining predictive models of discourse functions. USP 7,542,903, 06-2-2009

Systems and Methods for Dynamic Reading Fluency Instruction and Improvement. USP 7,455,522, 11-25-2008

Systems and Methods for Determining and Using Interaction Models. USP 7,415,414, 08-19-2008

Dynamically changing the levels of reading assistance and instruction to support the needs of different individuals. USP 7,386,453, 06-10-2008

System and Method for Summarization Combining Natural Language Generation with Structural Analysis. USP 7,305,336, 12-4-2007

Systems and Methods for Resolving Ambiguity. USP 7,283,958, 10-16-2007

System and method for Generating Analytic Summaries. USP 7,092,872, 08-15-2006

System and method for teaching writing using microanalysis of text. USP 7,013,259, 03-14-2006

System and method for writing analysis using the linguistic discourse model, USP 6,961,692, 11-1-2005

System and method for teaching second language writing skills using the linguistic discourse model. USP 6,866,510, 03-15-2005

## IBM

Method and System for distributed autonomous maintenance of bidirectional hyperlink metadata on the web and similar hypermedia repository. USP 7,266,762, 09-4-2007

System and Method for focussed web crawling. USP 6,418,433, 07- 9-2002

details at <http://mhvdberg.com/patents.html>

## Publications

### **A Semantic Account of Contextual Valence Shifting**

with: Livia Polanyi; in: Tracy Holloway King and Valeria de Paiva (eds.). From Quirky Case to Representing Space. Papers in the Honor of Annie Zaenen. CSLI Publications. p.195–208. 2013.

### **Discourse Structure and Sentiment**

with: Livia Polanyi; in IEEE 11th International Conference on Data Mining Workshops, pp.97-102, 2011.

### **LiveTree: An Integrated Workbench for Discourse Processing**

with: Gian Lorenzo Thione, Christopher Culy and Livia Polanyi; Proceedings of the ACL2004 Workshop on Discourse Annotation Barcelona, Spain July 25-26, 2004.

### **Sentential Structure and Discourse Parsing**

with: Livia Polanyi, Christopher Culy, Gian Lorenzo Thione and David Ahn. Proceedings of the ACL2004 Workshop on Discourse Annotation Barcelona, Spain, July 25-26, 2004.

### **Hybrid Text Summarization: Combining external relevance measures with Structural Analysis**

with: Livia Polanyi, Christopher Culy and Gian Lorenzo Thione; Proceedings of the ACL2004 Workshop Text Summarization Branches Out, Barcelona, Spain July 25-26, 2004

### **A Rule Based Approach to Discourse Parsing**

with: Livia Polanyi, Christopher Culy and Gian Lorenzo Thione; Proceedings of the 5th SIGdial Workshop in Discourse And Dialogue. Cambridge, MA USA, pp. 108-117. May 1, 2004

### **Discourse Structure and Sentential Information Structure: An Initial Proposal**

with: Livia Polanyi and David Ahn; in: the Journal of Logic, Language and Information 12-3:337-350 Kluwer Academic Publishers, Dordrecht, The Netherlands 2003

### **Making Ontologies Work for Resolving Redundancies Across Documents**

with: Livia Polanyi FXPAL, Dick Crouch, PARC/NLTT and Danny Bobrow, Cleo Condoravdi, John Everett, Valeria Paiva and Reinhard Stolle, PARC/RDC; in Communications of the ACM, February 2002, Vol. 45, Number 2: p.55–60, 2002.

### **Counting Concepts**

with: Dick Crouch and Cleo Condoravdi; in Robert van Rooy and Martin Stokhof (eds.), Proceedings of the Thirteenth Amsterdam Colloquium, ILLC/Department of Philosophy, University of Amsterdam, 2001.

### **Preventing Existence**

with: Dick Crouch, PARC/NLTT and Danny Bobrow, Cleo Condoravdi, John Everett, Valeria Paiva and Reinhard Stolle, PARC/RDC; in: Proceedings of the ACM Conference on Formal Ontology in Information Systems, 2001.

### **Questions as First Class Citizens**

in: Paul Dekker (ed), Proceedings of the Twelfth Amsterdam colloquium, ILLC/Department of Philosophy, University of Amsterdam, 1999.

### **Logical Structure and Discourse Anaphora Resolution**

with: Livia Polanyi; in: Dan Cristea, Nancy Ide, Daniel Marcu (eds.), Proceedings of the Workshop on The relation of Discourse/Dialogue Structure and Reference, 37th Annual Meeting of the Association of Computational Linguistics, 1999.

### **Distributed Hypertext Resource Discovery Through Examples**

with: Soumen Chakrabarti and Byron Dom; in Malcolm P. Atkinson, Maria E. Orlowska, Patrick Valduriez, Stanley B. Zdonik, Michael L. Brodie (e), VLDB'99, Proceedings of 25th International Conference on Very Large Data Bases, September 7-10, 1999, Edinburgh, Scotland, 375-386. Morgan Kaufman.

### **Focused Crawling: a New Approach to Topic-Specific Web Resource Discovery**

with: Soumen Chakrabarti and Byron Dom; in: *Computer Networks*, Volume 31, p. 1623--1640. Elsevier. 1999. Amsterdam, the Netherlands.

### **Some Aspects of the Internal Structure of Discourse. The Dynamics of Nominal Anaphora**

PhD-thesis, ILLC-DS-1996-3, Institute for Logic Language and Computation (ILLC), University of Amsterdam, 1996.

### **Dynamic Generalized Quantifiers**

in: van der Does, J. and van J. Eijck, (eds.), *Quantifiers, Logic, and Language*, CSLI Lecture Notes No.54, Stanford, California, 1996.

### **Discourse Grammar and Dynamic Logic**

in: Dekker, P. and M. Stokhof, (eds.): *Proceedings of the Tenth Amsterdam Colloquium*, ILLC/Department of Philosophy, University of Amsterdam, 1996.

### **Discourse Structure and Discourse Interpretation**

with: L. Polanyi; in Dekker, P. and M. Stokhof, (eds.): *Proceedings of the Tenth Amsterdam Colloquium*, ILLC/Department of Philosophy, University of Amsterdam, 1996.

### **Discourse Grammar and Verb Phrase Anaphora**

with: H. Prüst and R. J. H. Scha; *Linguistics and Philosophy* 17, 261--327 1994.

### **Plurality**

in: R.E. Asher and J.M.Y. Simpson (eds.) *The Encyclopedia of Language and Linguistics*, Pergamon Press. 3198--3200, 1994.

### **A Corpus-based Approach to Semantic Interpretation**

with: R. Bod and R. Scha; in Dekker, P. and M. Stokhof, (eds.): *Proceedings of the Ninth Amsterdam Colloquium*, ILLC/Department of Philosophy, University of Amsterdam, 1994.

### **A Direct Definition of Generalized Dynamic Quantifiers**

in Dekker, P. and M. Stokhof, (eds.): *Proceedings of the Ninth Amsterdam Colloquium*, ILLC/Department of Philosophy, University of Amsterdam, 1994.

### **Full Dynamic Plural Logic**

in: Bimbó, K. and A. Máté, (eds.), *Proceedings of the Fourth Symposium on Logic and Language*, Budapest, 1993.

### **Dynamic semantics**

in: Pinkal, M., R. Scha, and L. Schubert, (eds.), *Semantic Formalisms in Natural Language Processing*, No.57 in *Dagstuhl-Seminar-Report*, Schloß Dagstuhl, Wadern, Germany, 1993.

### **A formal grammar tackling verb-phrase anaphora**

with: H. Prüst and Remko Scha; *Technical Report CL-91-03*, ILLC, Amsterdam, 1991.

### **A Coherent Approach to Underspecification in Natural Language Discourse**

with: H. Prüst; in: Caenepeel M. et.al. (eds.), *Workshop on Discourse Coherence*, University of Edinburgh, 1991.

### **Common Denominators and Default Unification**

with: H. Prüst; in: T. van der Woude and W. Sijtsma (eds.), *Computational Linguistics in the Netherlands. Papers for the first CLIN meeting*, OTS-WP-CL-91-001 in *OTS Working Papers*, University of Utrecht, 1990.

### **A Dynamic Logic for Plurals**

in: Stokhof, M. and L. Torenvliet (eds.), *Proceedings of the Seventh Amsterdam Colloquium*, ILLI, University of Amsterdam, 1990.