

Lexical and Statistical Semantics in Professional Search

Allan Hanbury



FAKULTÄT
FÜR INFORMATIK
Faculty of Informatics



COMPLEXITY
SCIENCE
HUB
VIENNA

Google

Österreich

|

Google Search

I'm Feeling Lucky



Bing

|



DuckDuckGo

|





Attorney Docket No. APL1P223/P2727

PATENT APPLICATION

FOR

METHOD AND APPARATUS FOR ACCELERATED SCROLLING

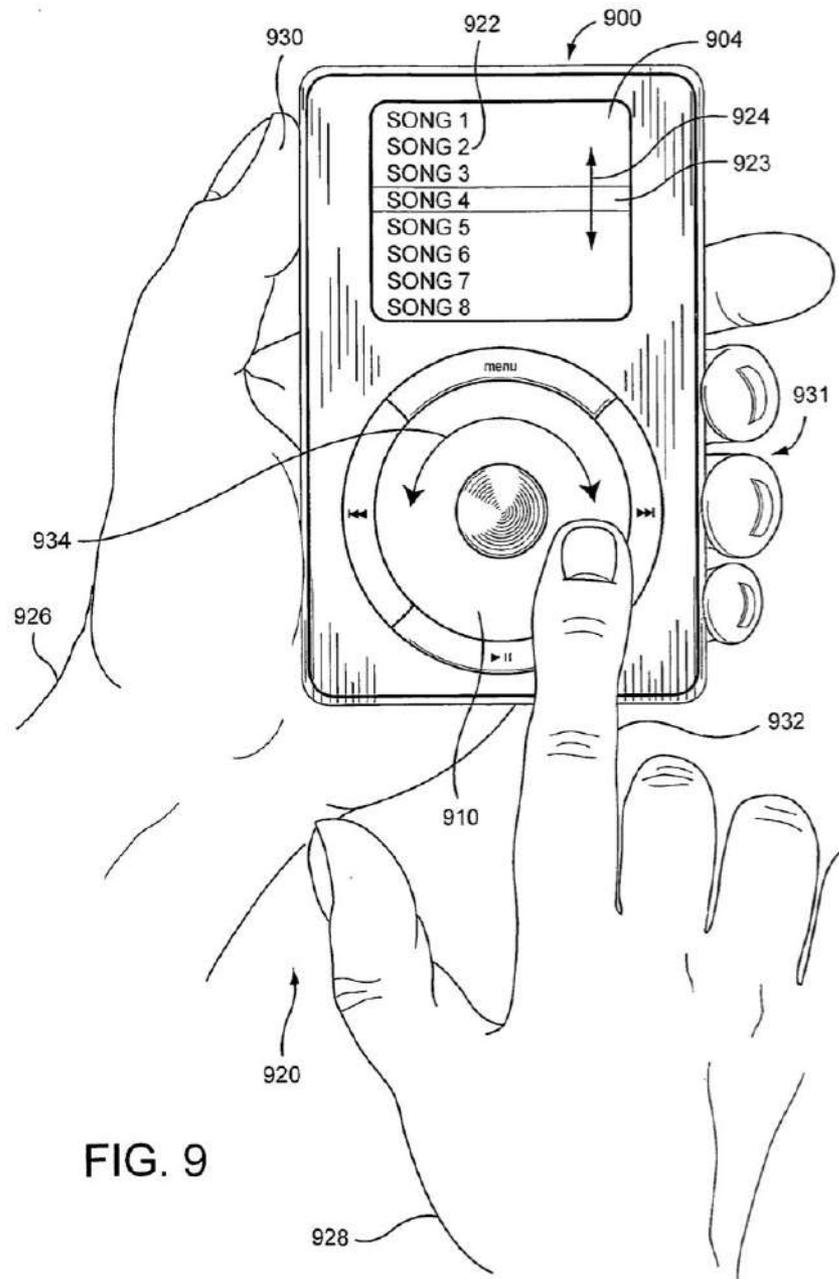
Inventors: 1. Robert Tsuk
2. Jeffrey L. Robbin

Assignee: Apple Computer, Inc.

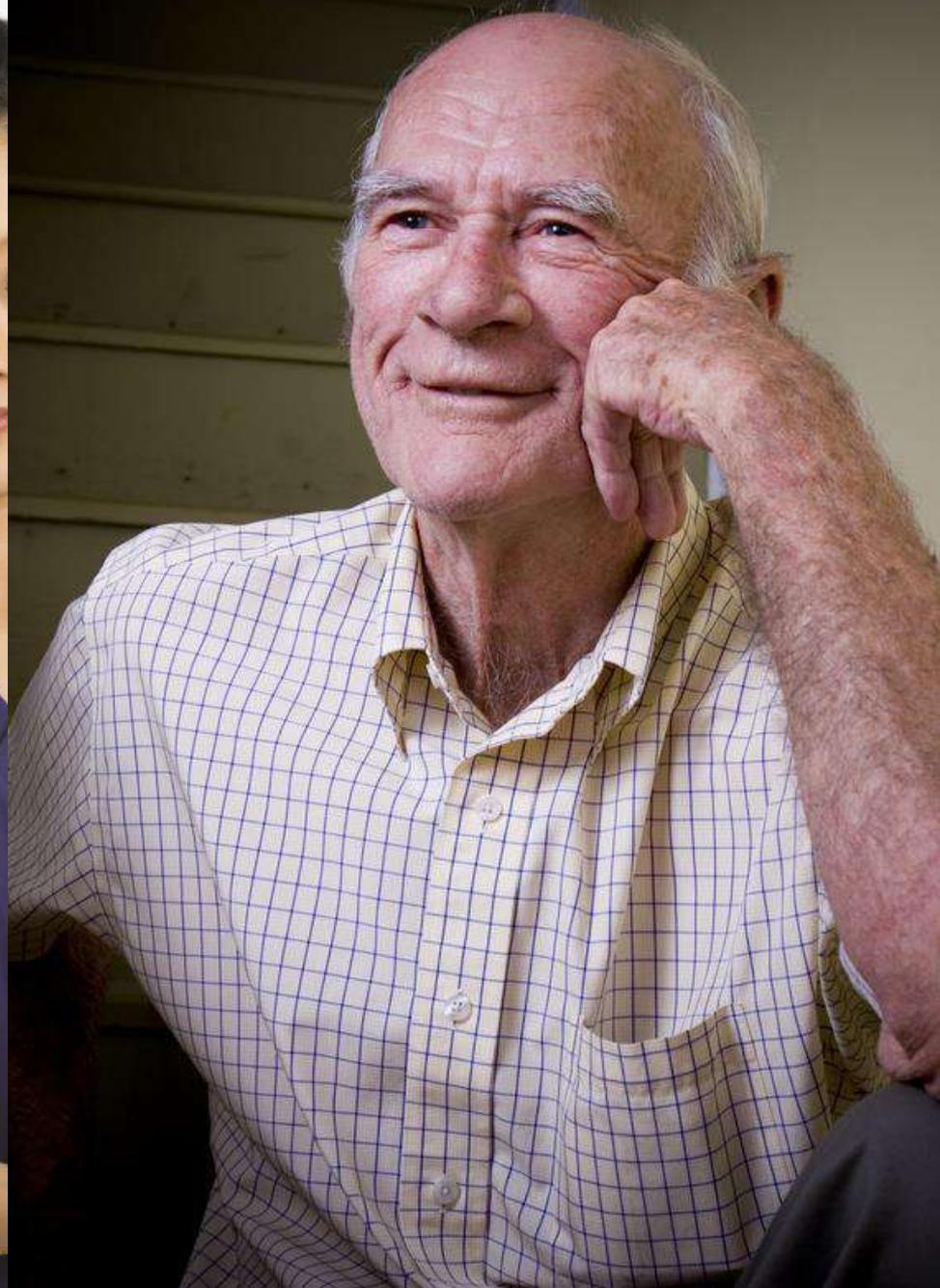
BEYER WEAVER & THOMAS, LLP
(650) 961-8300

Claims

1. A method for scrolling through portions of a data set, said method comprising: receiving a number of units associated with a rotational user input; determining an acceleration factor pertaining to the rotational user input; modifying the number of units by the acceleration factor; determining a next portion of the data set based on the modified number of units; and presenting the next portion of the data set.
2. A method as recited in claim 1, wherein the data set pertains to a list of items, and the portions of the data set include one or more of the items.
3. A method as recited in claim 1, wherein the data set pertains to a media file, and the portions of the data set pertain to one or more sections of the media file.
4. A method as recited in claim 3, wherein the media file is an audio file.
5. A method as recited in claim 1, wherein the rotational user input is provided via a rotational input device.
6. ...







Search Systems



Context

- People
- Tasks
- Documents
- ...

Technology

- Natural Language Processing
- Data-driven approaches
- ...

Natural Language Processing

This is particularly preferable
in the case of liquid
injection molding (LIM).



This/**DT** is/**VBZ** particularly/**RB** preferable/**JJ**
in/**IN** the/**DT** case/**NN** of/**IN** liquid/**JJ**
injection/**NN** molding/**NN** (LIM/**NN**).

PIC Detection

Population

Intervention

Comparison

Input:

*The bioavailability of
nasogastric versus
trovafloxacin in healthy
subjects*

The screenshot shows the PubMed.gov search interface. At the top left is the PubMed logo and the text "US National Library of Medicine National Institutes of Health". To the right is a search box containing "PubMed" and a dropdown arrow, followed by an empty search input field and a link to "Advanced". Below the search bar, the "Format" is set to "Abstract". The search results display the journal citation "Am J Surg. 1998 Dec;176(6A Suppl):23S-26S." followed by the title "The bioavailability of nasogastric versus" and the authors "Vincent J¹, Teng R, Pelletier SM, Willavize SA, Friedman HL."

PubMed.gov
US National Library of Medicine
National Institutes of Health

PubMed

Advanced

Format: Abstract

Am J Surg. 1998 Dec;176(6A Suppl):23S-26S.

The bioavailability of nasogastric versus

Vincent J¹, Teng R, Pelletier SM, Willavize SA, Friedman HL.

PIC Detection

Population

Intervention

Comparison

Input:

The bioavailability of nasogastric versus trovafloxacin in healthy subjects

Algorithm

Pattern 1 (of I versus C in P)

MATCH

Pattern 2 (...)

NO MATCH

⋮

⋮

Pattern N (...)

NO MATCH

Output:

P: healthy subjects

I: nasogastric

C: trovafloxacin

Title: Effects of metformin on the body composition in subjects with risk factors for type 2 diabetes.

Abstract:

OBJECTIVE

To measure the effect of metformin on the body composition, insulin resistance and sensitivity in subjects with risk factors for type 2 diabetes mellitus (type 2 DM).

METHODS

Placebo-controlled clinical trial. Twenty-three subjects with risk factors for type 2 DM were randomly assigned to receive 850 mg of metformin or a placebo twice a months. Before and after the treatment, the body mass index and waist/hip ratio were calculated, the body composition was measured through bioelectric impedance of blood glucose, insulin, triglycerides and cholesterol were measured. The level of insulin resistance was calculated by the homeostatic model and the level of sensitivity quantitative insulin sensitivity check index method. The Wilcoxon rank test was used.

RESULTS

Twenty-one subjects completed the study, 12 of the metformin group and nine of the placebo group. In the metformin group, there was a decrease in fat weight from 9.2 kg, $p < 0.01$, an increase in lean weight from 57.05 +/- 13.6 to 61.9 +/- 16.5 kg, $p < 0.01$, an increase in basal metabolism from 1735 +/- 413 to 1878 +/- 505 cal, an increase in body water, $p < 0.05$. There was no significant decrease in insulin resistance. In the placebo group, the blood glucose increased from 84.7 +/- 13 to 96.05. There were no significant modifications in lipids.

CONCLUSIONS

The administration of metformin for 2 months improves the parameters of body composition and insulin dynamics in subjects with risk factors for type 2 DM.

Sentiment:

Positive: The administration of metformin for 2 months improves the parameters of body composition and insulin dynamics in subjects with risk factors for type 2 DM.

Sample Size: 23



EXPERT OPINION

Contact Editor: **Brian Brannon**, bbrannon@computer.org

The Unreasonable Effectiveness of Data

Alon Halevy, Peter Norvig, and Fernando Pereira, *Google*

2009

Data-Driven

$$P(\text{Word}_1 \mid \text{Word}_2)$$



“Every time I fire a linguist,
the performance of the
speech recognizer goes up.”

Frederick Jelinek, IBM

Around 1985



Yoav Goldberg

[Follow](#)

Senior Lecturer at Bar Ilan University. Working on NLP. Recently with Neural Nets. Published a book about...

Jun 8 · 14 min read

An Adversarial Review of “Adversarial Generation of Natural Language”

Or, for f••ks sake, DL people, leave language alone and stop saying you solve it.



“In most cases, the meaning
of a word is its use.”

*Ludwig Wittgenstein,
Philosophical
Investigations (1953)*

A bottle of **tesgüino** is on the table.

Everybody likes **tesgüino**.

Tesgüino makes you drunk.

We make **tesgüino** out of corn.

Statistical semantics extracts
lexical information from the
statistics of large amounts of
text

Word Embedding

Source Text

Training Samples

The quick brown fox jumps over the lazy dog. →

(the, quick)
(the, brown)

The quick brown fox jumps over the lazy dog. →

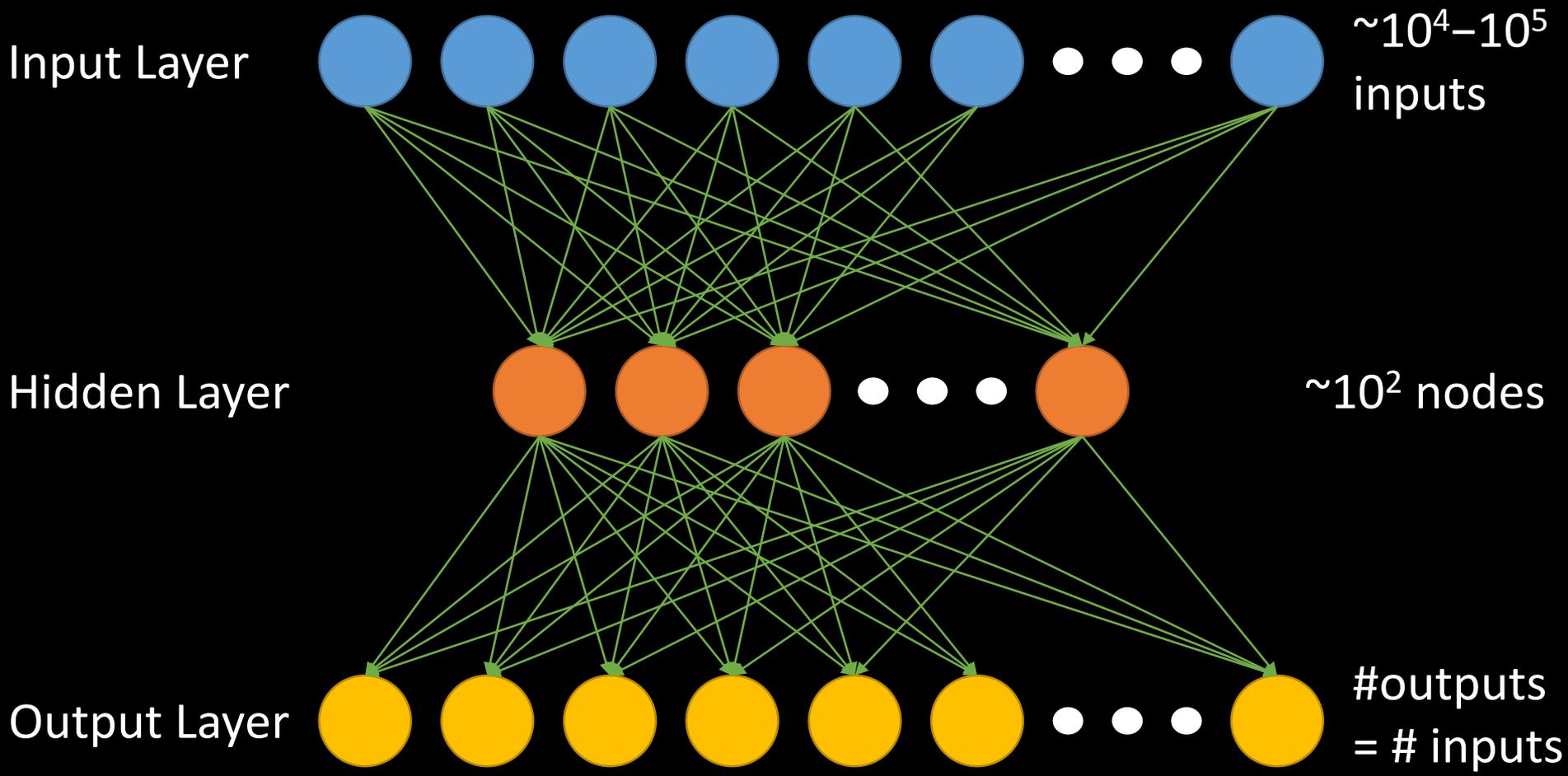
(quick, the)
(quick, brown)
(quick, fox)

The quick brown fox jumps over the lazy dog. →

(brown, the)
(brown, quick)
(brown, fox)
(brown, jumps)

The quick brown fox jumps over the lazy dog. →

(fox, quick)
(fox, brown)
(fox, jumps)
(fox, over)



What is more similar to a typewriter:

(a) a bird, or

(b) a snowball?

Similarity Threshold

	book		dwarfish
	books	0.82	corpulent 0.44
	foreword	0.77	hideous 0.43
	author	0.74	unintelligent 0.42
Threshold	published	0.73	wizened 0.42
	preface	0.69	catoblepas 0.42
Top k	republished	0.68	creature 0.42
	reprinted	0.68	humanoid 0.41
	afterword	0.67	grotesquely 0.41
	memoir	0.67	tomtar 0.41

Navid Rekabsaz, Mihai Lupu, Allan Hanbury, Guido Zuccon, Exploration of a Threshold for Similarity based on Uncertainty in Word Embedding. In Proceedings of the European Conference on Information Retrieval Research (ECIR 2017)

Navid Rekabsaz, Mihai Lupu, Allan Hanbury, Uncertainty in Neural Network Word Embedding: Exploration of Threshold for Similarity. Proc. Neu-IR Workshop of the ACM Conference on Research and Development in Information Retrieval (NeuIR-SIGIR 2016)

Word Embedding in Search

knowledge

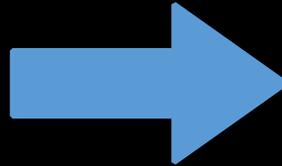


knowledge

wisdom

knowledge

understanding

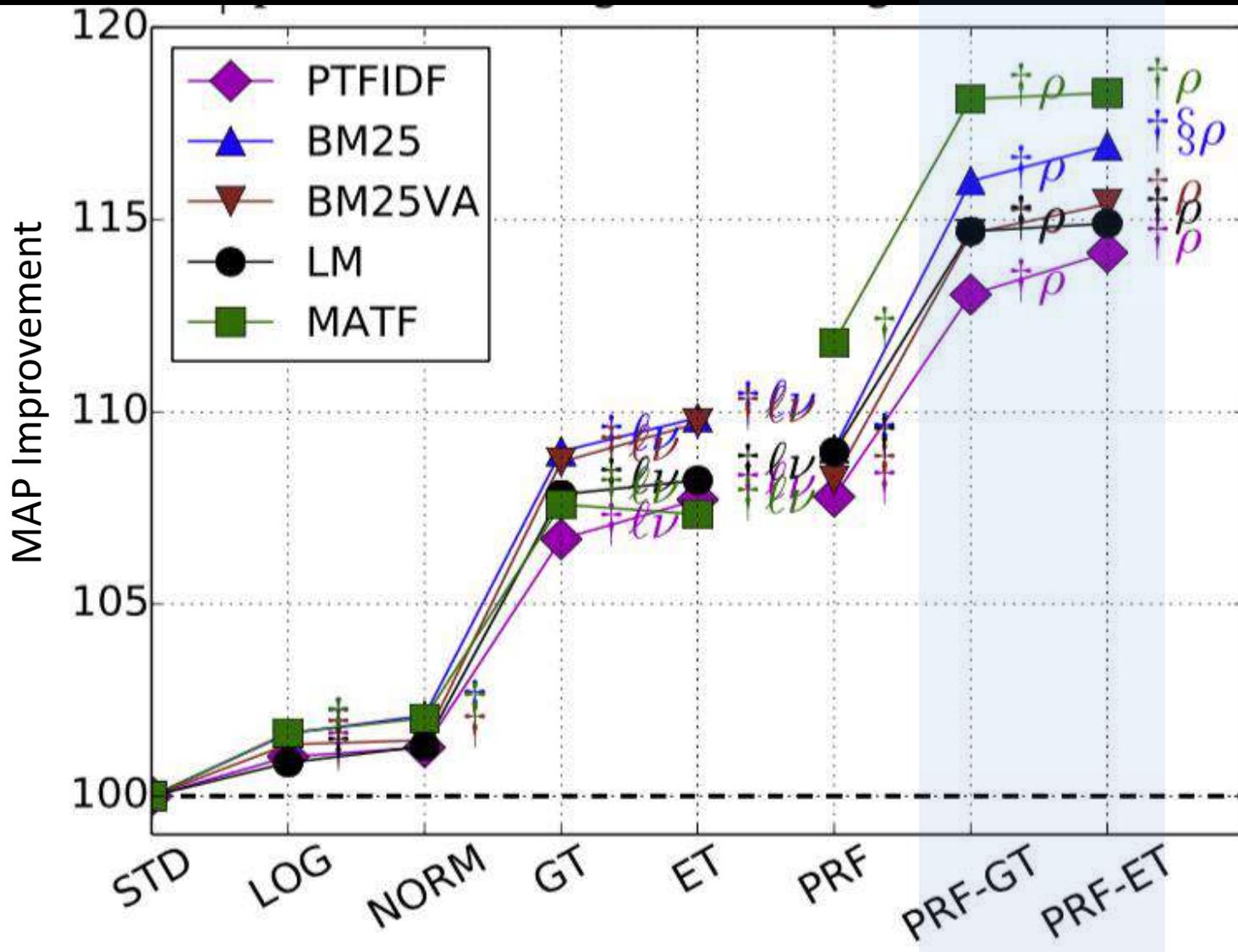


knowledge

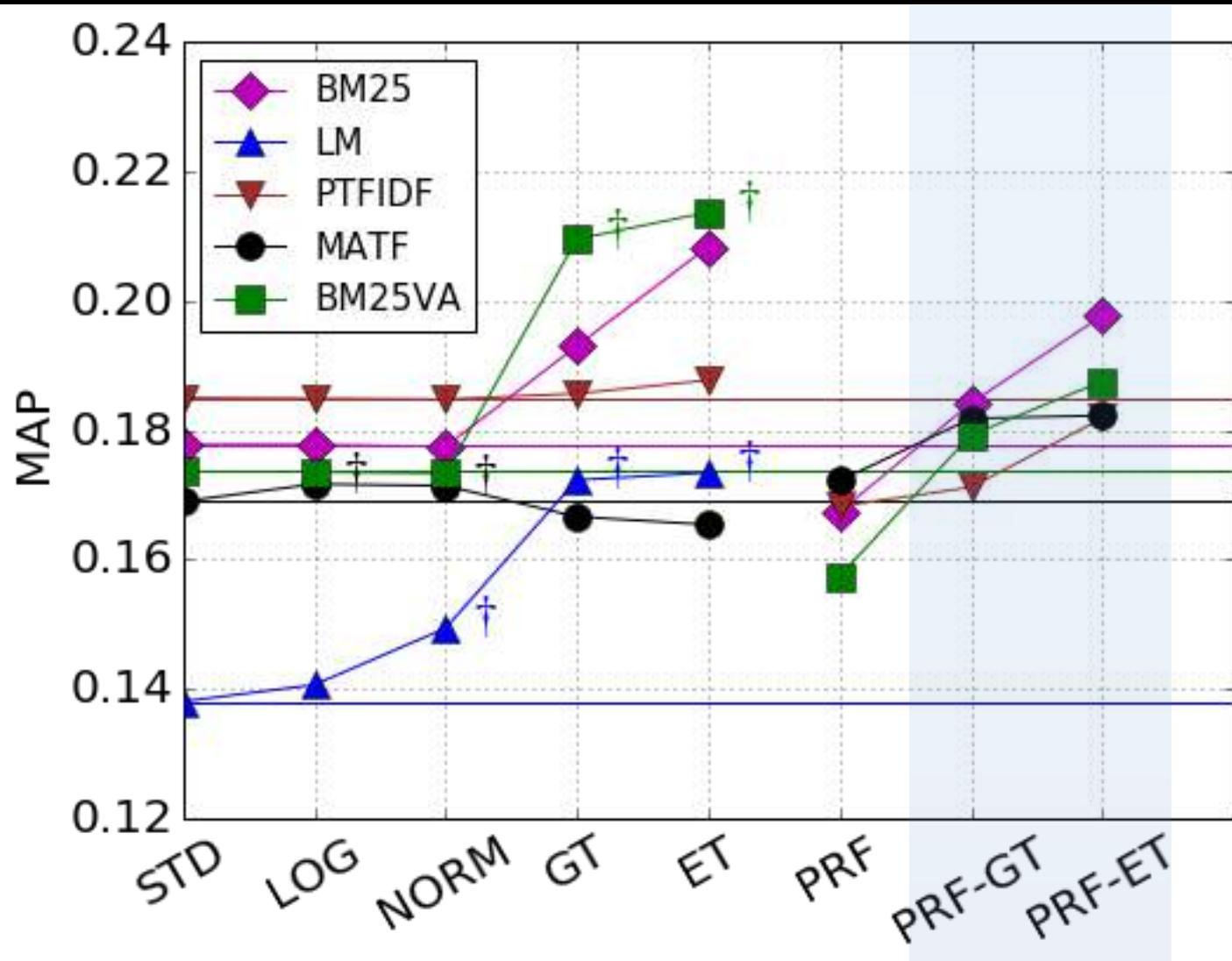
$0.7 \times \textit{knowledge}$

knowledge

$0.8 \times \textit{knowledge}$



Results for Patent Search



Document Length

Sentence Complexity

New Terminology

Multi-Word Terms in Patents

coating method

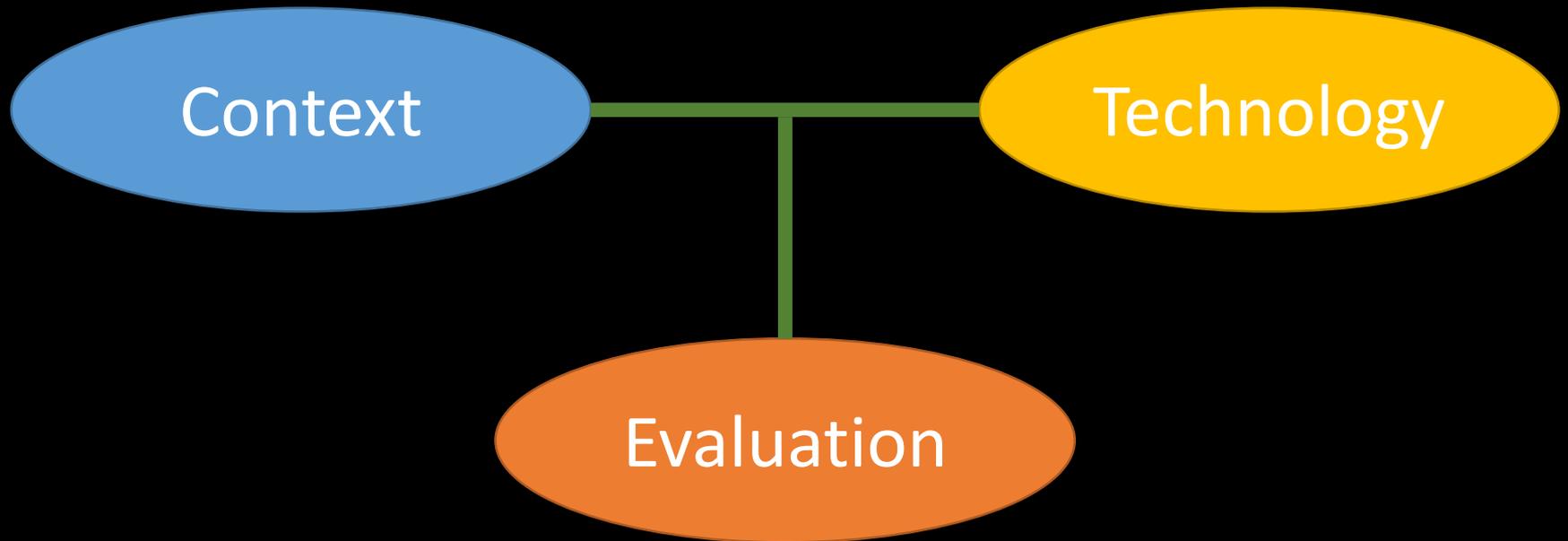
memory data processor

complex programmable logic device

rear cross frame member

document editing device

Search Systems



Acknowledgements

Linda Andersson, Alexandros Bampoulidis,
Tobias Fink, Aldo Lipani, Mihai Lupu,
João Palotti, Florina Piroi, Navid Rekabsaz,
Abdel Aziz Taha, Markus Zlabinger



FFG



SELF OPTIMIZER





FAKULTÄT
FÜR INFORMATIK

Faculty of Informatics



COMPLEXITY
SCIENCE
HUB
VIENNA

Univ.-Prof. Dr. Allan Hanbury

Institute for Information Systems Engineering
TU Wien
Favoritenstraße 9-11/194-04
1040 Vienna
Austria

Telephone: +43 1 58801 188310

Mobile: +43 676 978 0991

e-Mail: allan.hanbury@tuwien.ac.at

