

# Thaer Samar

---

## Contact Information:

**Mobile Phone** +972 (0) 59 4676 980  
**Email** thsammar@gmail.com  
**LinkedIn** <https://nl.linkedin.com/in/thaersamar>

## Education

**Apr 2012 - Sep 2017**

**Centrum Wiskunde & Informatica, The Netherlands**

PhD Candidate, PhD thesis was defended in Radboud University Nijmegen, The Netherlands on 30 October 2017. Promotor: Prof. dr. ir. Arjen P. de Vries

My research focuses on accessing and analyzing large scale Web data over time (e.g. Web Archives). During this period I was a member in the Web Archive Retrieval Tools (WebART) project. Project Leader: dhr. dr. ir. J. (Jaap) Kamps

**Oct 2009 - Dec 2011**

**Saarland University, Germany**

M.Sc. in Computer Science

Grade - 1.4 (A-)

**Master Thesis**

Title: Scalable Distributed Time-Travel Text Search

Grade: 1.3

Advisor: Professor Klaus Berberich

**Sep 2002 - Jun 2007**

**Al-Quds University, Palestine**

B.Eng. in Computer Engineering

Grade - 87.6%, Rank: first

## Experience

**Sep 2018 - Now**

**Palestine Technical University, Palestine**

Assistant Professor

**Sep 2016 - Aug 2018**

**Geophy, The Netherlands**

Senior Data Engineer/Scientist

**Apr 2016 - Sep 2016**

**Centrum Wiskunde & Informatica, The Netherlands**

Researcher in the Information Access Group

**Apr 2012 - Aug 2016**

**Centrum Wiskunde & Informatica, The Netherlands**

Scientific Programmer - PhD Candidate

I was involved in the Web Archive Retrieval Tools (WebART) project. WebART is an interdisciplinary research project funded by NWO. Its team combines expertise in Computer Science, Information Science and New Media research from the University of Amsterdam and the Centrum for Wiskunde and Informatica. The cultural heritage partner of the project is the National Library of the Netherlands (*KB*). My role in the project was to use my expertise as a data scientist and big data engineer (analytic) to access the Dutch Web archive collections at scale.

Sep 2007 - Jun 2009	<b>Computer Systems Engineering, Berzeit University, Palestine</b> Teaching Assistant
Jun 2007 - Aug 2007	<b>Computer Media Center, Palestine</b> Computer and Network maintenance
Jul 2006 - Sep 2006	<b>Palestine Cellular Communication (LTD)- Jawwal, Palestine</b> Network Engineering (internship)

## Awards

**German Academic Exchange Service (DAAD)**  
Master's scholarship (two years)

**The Ministry of Higher Education Palestine**  
Honoring the first ranked students in Palestinian Universities

**Al-Quds University**  
Honor list (nine semesters out of ten)

## Languages

<b>Arabic</b>	Native language
<b>English</b>	Fluent in speaking, writing and reading (full professional proficiency)
<b>German</b>	Professional working proficiency (B1 level)
<b>Dutch</b>	Basics

## Technical Skills

<b>Programming</b>	Java, C, C++, SQL, Python
<b>Tools</b>	Hadoop/MapReduce, Hadoop/HDFS, Pig Latin, HBase, Spark, Lucene, Solr, Terrier, Indri, Weka, Latex, R, Git, MonetDB, Kafka, RDF, SPARQL

## Publications

### Journal Articles

- 2017 **Thaer Samar** and Myriam C. Traub and Jacco van Ossenbruggen and Lynda Hardman and Arjen P. de Vries. *Quantifying retrieval bias in Web archive search*. International Journal on Digital Libraries.
- 2015 **Thaer Samar** and Alejandro Bellogin and Arjen P. de Vries. *The Strange Case of Reproducibility vs. Representativeness in Contextual Suggestion Test Collections*. Information Retrieval Journal.
- 2015 Hugo C. Huurdeman and Jaap Kamps and **Thaer Samar** and Arjen P. de Vries and Anat Ben-David and Richard A. Rogers. *Lost but not forgotten: finding pages on the unarchived web*. International Journal on Digital Libraries.

### Conference Proceedings

- 2016 **Thaer Samar** and Myriam C. Traub and Jacco van Ossenbruggen and Arjen P. de Vries.

*Comparing Topic Coverage in Breadth-First and Depth-First Crawls Using Anchor Texts.*  
The 20th International Conference on Theory and Practice of Digital Libraries  
TPDL 2016, Germany.

- 2016 Myriam C. Traub and **Thaer Samar** and Jacco van Ossenbruggen and Jiyin He and Arjen P. de Vries and Lynda Hardman: *Querylog-based Assessment of Retrievability Bias in a Large Newspaper Corpus.*  
The 16th ACM/IEEE-CS on Joint Conference on Digital Libraries JCDL 2016, USA.
- 2014 **Thaer Samar** and Hugo C. Huurdeman and Anat Ben-David and Jaap Kamps and Arjen P. de Vries. *Uncovering the unarchived web.* The 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '14, Australia.
- 2014 Alejandro Bellogin and **Thaer Samar** and Arjen P. de Vries and Alan Said. *Challenges on Combining Open Web and Dataset Evaluation Results: The Case of the Contextual Suggestion Track.* Advances in Information Retrieval - Proceedings of the 36th European Conference on IR Research, ECIR 2014, The Netherlands.
- 2014 Hannes Mühleisen and **Thaer Samar** and Jimmy Lin and Arjen P. de Vries. *Column Stores as an IR Prototyping Tool.* Advances in Information Retrieval - Proceedings of 36th European Conference on IR Research, ECIR 2014, The Netherlands.
- 2014 Hannes Mühleisen and **Thaer Samar** and Jimmy Lin and Arjen P. de Vries. *Old dogs are great at new tricks: column stores for ir prototyping.* The 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '14, Australia.
- 2014 Hugo C. Huurdeman and Anat Ben-David and Jaap Kamps and **Thaer Samar** and Arjen P. de Vries. *Finding pages on the unarchived Web.* IEEE/ACM Joint Conference on Digital Libraries, JCDL 2014, United Kingdom.
- 2013 Hugo C. Huurdeman and Anat Ben-David and **Thaer Samar**. *Sprint methods for web archive research.* Proceedings of Web Science 2013 (WebSci '13), France.

### Refereed Workshop Proceedings

- 2015 **Thaer Samar** and Arjen P. de Vries. *Temporal Anchor Text as Proxy for Real User Queries.* Proceedings of the 5th International Workshop on Semantic Digital Archives (SDA 2015) co-located with the International Conference on Theory and Practice of Digital Libraries, TPDL 2015, Poland.
- 2015 **Thaer Samar** and Alejandro Bellogin and Arjen P. de Vries. *Improving Contextual Suggestions using Open Web Domain Knowledge.* Proceedings of the International Workshop on Social Personalisation & Search, SPS 2015, co-located with the 38th Annual ACM SIGIR Conference (SIGIR 2015), Chile, 2015.

### Working Notes

- 2014 **Thaer Samar** and Arjen P. de Vries and Alejandro Bellogin. *Better Contextual Suggestions in ClueWeb12 Using Domain Knowledge Inferred from The Open Web.* Proceedings of The Twenty-Third Text REtrieval Conference, TREC 2014, USA.
- 2013 Alejandro Bellogin and Gebrekirstos G. Gebremeskel and Jiyin He and Alan Said and **Thaer Samar** and Arjen P. de Vries and Jimmy Lin and Jeroen B. P. Vuurens. *CWI and TU Delft Notebook TREC 2013: Contextual Suggestion, Federated Web Search, KBA, and Web Tracks.* Proceedings of The Twenty-Second Text REtrieval Conference, TREC 2013, USA.

