

WikipediaSearch

An application of Personalized PageRank on a Wikipedia subset

Roberto Corti

Information Retrieval exam

December 6, 2020



UNIVERSITÀ
DEGLI STUDI DI TRIESTE



DATA SCIENCE &
SCIENTIFIC COMPUTING

Outline

Introduction

Wikipedia dataset

WikipediaSearch implementation

WikipediaSearch evaluation

Conclusion

Outline

Introduction

Wikipedia dataset

WikipediaSearch implementation

WikipediaSearch evaluation

Conclusion

WikipediaSearch

A brief introduction

WikipediaSearch is a user-interactive tool that computes a Personalized (or Topic Specific) PageRank over a Wikipedia corpus.

WikipediaSearch

A brief introduction

WikipediaSearch

Result for Roberto Corti

- [Mean_\(statistics\)](#)
- [Statistics](#)
- [Computer](#)
- [Mathematics](#)
- [Science](#)
- [Computer_science](#)
- [Message_\(computer_science\)](#)
- [Mean](#)
- [Sampling_\(statistics\)](#)
- [Physics](#)
- [Algorithm](#)
- [Error](#)
- [Object-oriented_programming](#)
- [Football_\(soccer\)](#)
- [Finance](#)
- [United_States_0944](#)
- [Sampling](#)
- [Number](#)
- [Computer_program](#)
- [People](#)
- [Programmer](#)

Output page: a rank of Wikipedia articles in which the user would be interested to read

The problem

"Bringing Order to the Web"

A recap of PageRank theory

From L.Page and S.Brin (1998)

Outline

Introduction

Wikipedia dataset

WikipediaSearch implementation

WikipediaSearch evaluation

Conclusion

Wikipedia ...

content...

Outline

Introduction

Wikipedia dataset

WikipediaSearch implementation

WikipediaSearch evaluation

Conclusion

PageRank implementation

Outline

Introduction

Wikipedia dataset

WikipediaSearch implementation

WikipediaSearch evaluation

Conclusion

Standard PageRank

Outline

Introduction

Wikipedia dataset

WikipediaSearch implementation

WikipediaSearch evaluation

Conclusion

Conclusion