An application of Personalized PageRank on a Wikipedia subset

Roberto Corti

Information Retrieval exam

December 6, 2020







Introduction

Wikipedia dataset

WikipediaSearch implementation

WikipediaSearch evaluation

### Introduction

Wikipedia datase

WikipediaSearch implementation

WikipediaSearch evaluation

A brief introduction

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

### A brief introduction

Introduction

```
lame: Roberto
 Surname: Cortl
Melcome to WikipediaSearch!
This demo performs a Personalized PageRank into the simple Wikipedia dump of April 2007 (https://dumps.wikimedia.org/other/static_html_dumps/April_2007/
simple) matching some user-defined topic interests. The final result will be a static html page that contains the rank of the Wikipedia pages according
to the topics selected by the user.
Please Roberto Corti, list your 5 main interests and rate each of them from 1 to 5
Topic 1: football
Rate football (lower=1, higher=5): 2
Topic 2: science
Rate science (lower=1, higher=5): S
Topic 3: physics
Rate physics (lower=1, higher=5): 3
 opic 4: statistics
Rate statistics (lower=1, higher=5): 3
 Topic 5: programming
      programming (lower=1, higher=5): 4
```

**Input interface**: user specifies the topics in which he/she has more interest



### A brief introduction

Introduction 0000

#### 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 (somer) • Finance - United States 09d4 Sampling • Number · Computer\_program • People • Programmer

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



# The problem

Introduction

"Bringing Order to the Web"



Introduction

# A recap of PageRank theory

From L.Page and S.Brin (1998)

Wikipedia dataset

WikipediaSearch evaluation

content...

WikipediaSearch implementation

WikipediaSearch evaluation

# PageRank implementation



Introduction

Wikipedia datase

WikipediaSearch implementation

WikipediaSearch evaluation

# Standard PageRank

Introduction

Wikipedia datase

WikipediaSearch implementation

WikipediaSearch evaluation

