Travel Search

Extreme Search

Ngoni Choga
Dylan Henderson
Shuaib Parker
Luqmaan Salie

Supervisor: Dr Maria Keet
Second reader: Assoc. Prof. Patrick Marais

Booking Flights

book a flight

from:* O.R. Tambo (Jo'burg) ▼ to:* Cape Town ▼

○ one-way ◉ return ○ multi city

departing: *	returning: *	adults	children	infants
24 May 2015	24 May 2015	1 ▼	0 ▼	0 ▼
		>12 years	2-11 years	<23 mnths

preferred class:

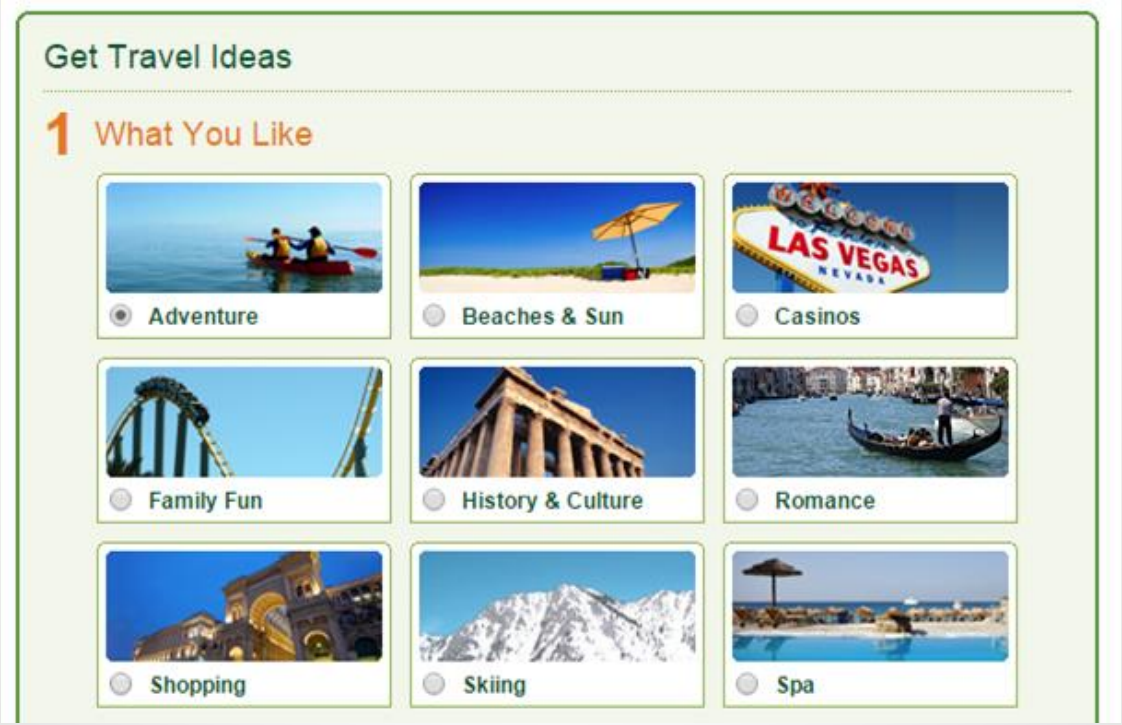
Business class availability is airline dependent.

economy ▼

All fields marked with * are required

search

Choosing Destinations



“I would like to go somewhere romantic for between R2000 - R10000”

- Mozambique

- Tanzania

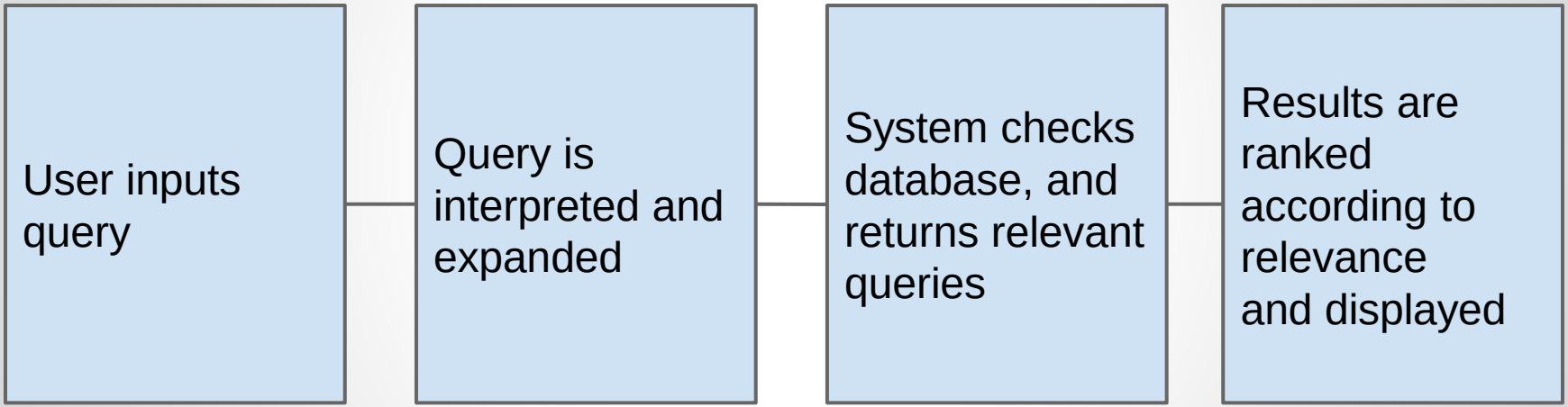
- Morocco

Project Methodology

Agile

- Increases customer feedback with iterations
- Provide better feature flexibility
- More input on deliverables from client

Search Process



Tasks

User interface design - Ngoni
Creation of faceted database - Dylan
Query formulation and expansion - Luqmaan
Relevance ranking - Shuaib

User Interface

HCI practices

- Choose user population
- Understand the users
- Prototypes
- User centered design
- Constant user involvement
- **Testing**

Testing

Assessment

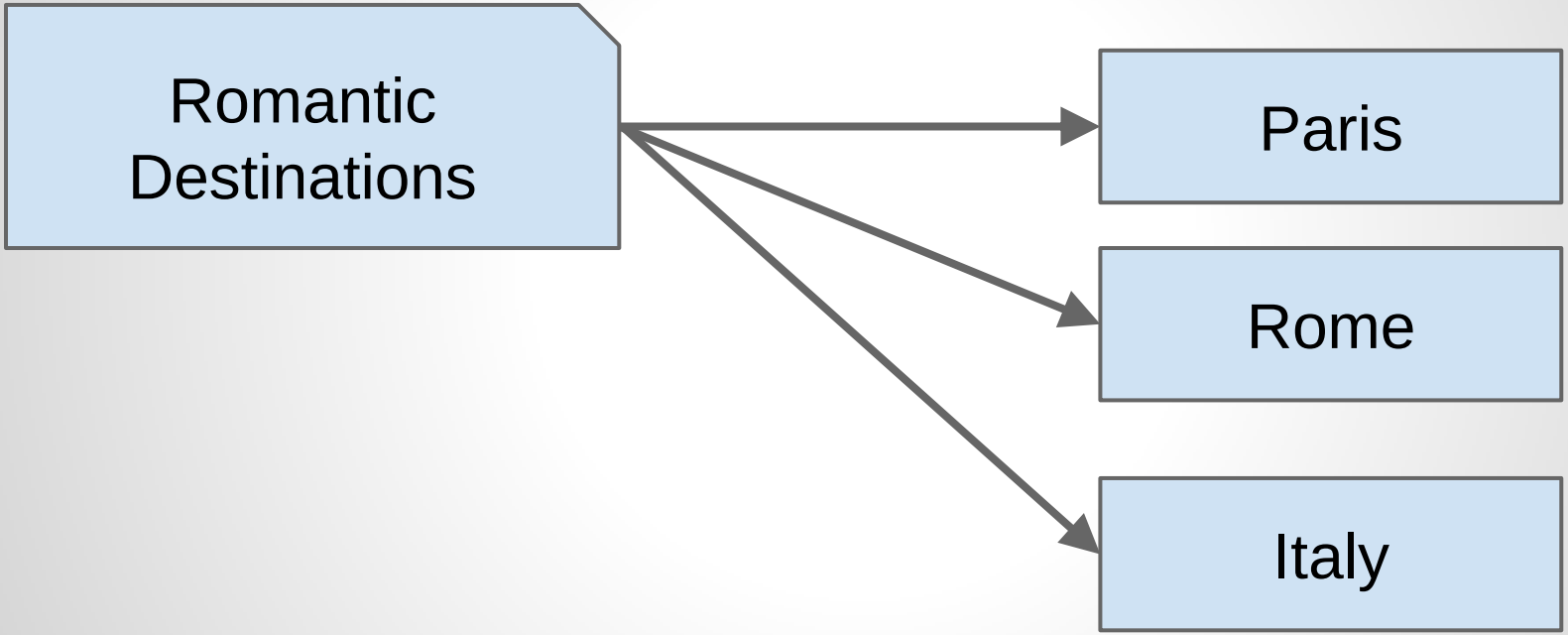
Automated - Google analytics
Empirical - Real users
Inspection - Heuristics, skills and experience of users

Users

Two main target user personae

- Frequent fliers (international students)
- Infrequent fliers

Faceted Search System



Basic Storage

Destinations

```
{
    Destination: France
    ImageUrl: http:/Image/Paris
    Categories:[
        - Romantic
        - Adventure
    ]
}
```

Indexing

```
{
    Category: Romantic
    Destinations:[
        - France
        - Rome
        - Belgium
    ]
}
```

Accessing Information

Travelstart API
TripAdvisor API
DBPedia
LOD (Linking Open Data) Cloud

Dynamic Gathering

Users Describe Destinations

-“Exciting”

-“Fresh”

Facets Updated

{

Category: Exciting

Destinations: [

-Paris

]

}

Project Plan and Evaluation

- Gathering, storing, accessing, and automatic updating
- User testing, returning correct results from category search
- Evaluate facet change
- How often facets are searched

Query Formulation and Expansion

Query types

- Text-based queries
- Faceted queries

Importance

- Optimal query structure
- Broaden search results

Query Formulation and Expansion

Formulation methods

- Stemming, normalisation, tokenisation, removing stop words

Expansion methods

- Local context analysis
- Real-time expansion
- Semantic query expansion

Query Formulation and Expansion

Challenge

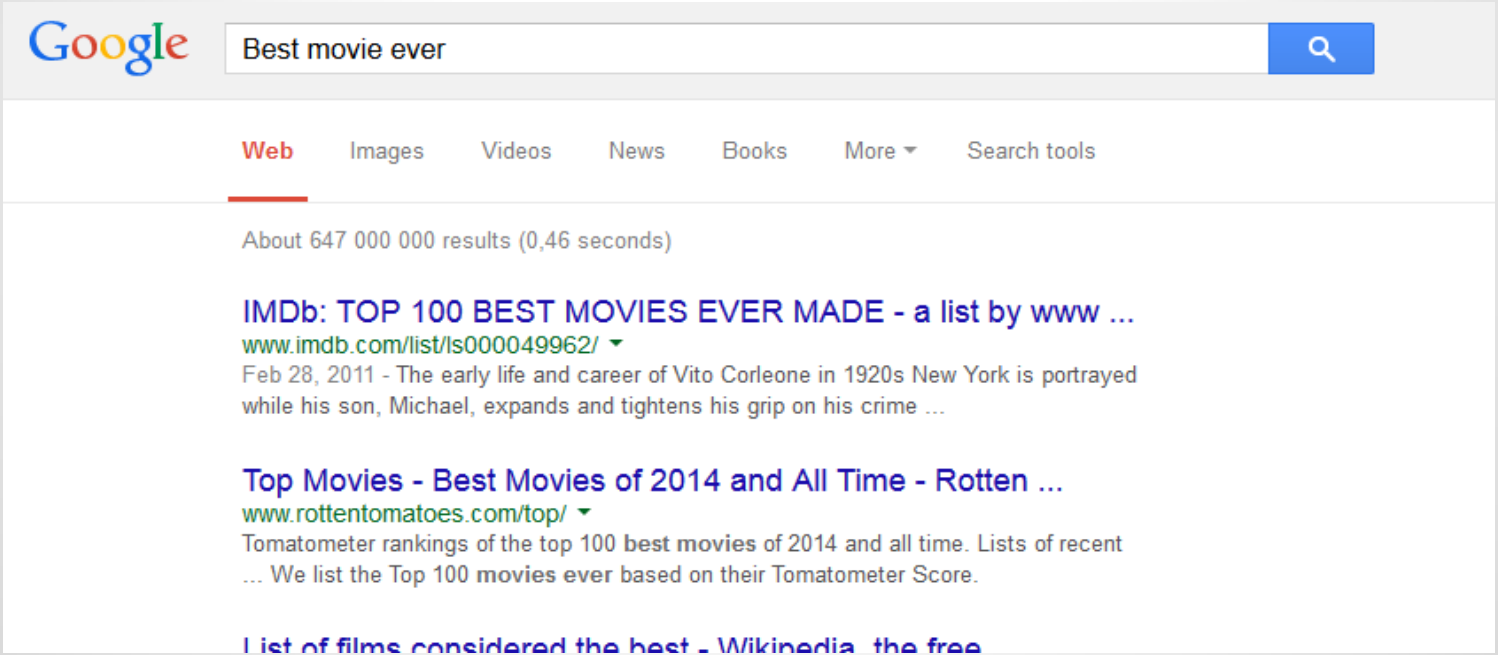
- Finding the optimal combination of formulation and expansion methods, terms, and speed

Evaluation methods

- User evaluation of expansion implementations
- F-measure - average between recall and precision
- Expansion term retrieval speed

Relevance Ranking

Users often look at only the first few results



Relevance Ranking

Category “weights” will determine rank

```
{
    Category: Romantic
    Destinations:[
        - Hawaii: 4
        - Rome: 2
        - Belgium: 1
    ]
}
```

```
{
    Category: Island
    Destinations:[
        - Hawaii: 5
        - Rome: 0
        - Belgium: 0
    ]
}
```

Relevance Ranking

Refine search results based on user activity

- Use tools such as Google analytics

Evaluation

- Discounted Cumulative Gain
- Mean average precision
- Speed tests - input to results

Overall System Evaluation

- Temporary integration
- “Try me” button
- User ratings
- **Integration with Travelstart**

Resources and Material

Front-end

- HTML
- CSS
- JavaScript

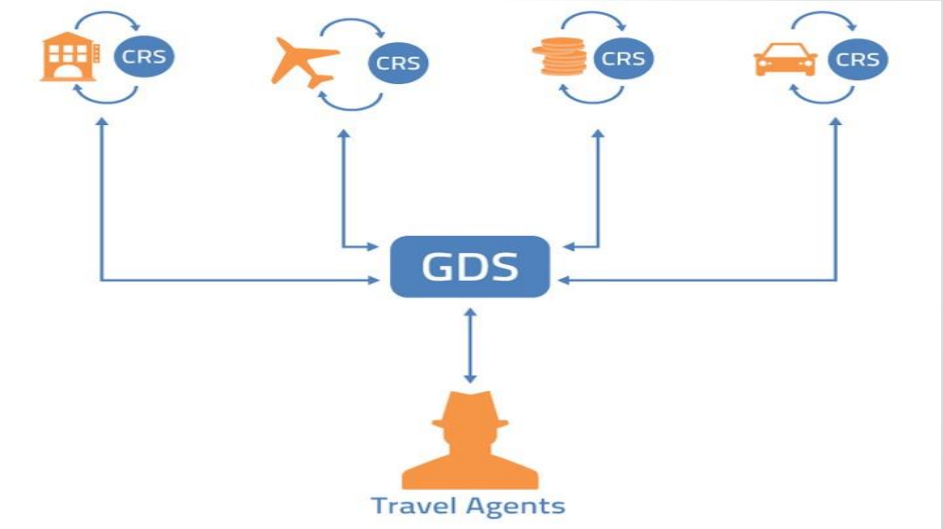
Back-end

- Java
- Node.js
- ASP.net
  section requirements

*Depends on

Additional Material

- Travelstart API via SOAP/REST
- Global Distribution Systems (GDS)
- DBpedia



Deliverables

- Project website
- Minimum viable product
- Second iteration
- Final system
- Report + poster + website

Outcomes

- Destination search system
- New method for finding destinations
- Competitive advantage for client

References

Raffaella Bernardi, Diego Calvanese, Luca Dini, Vittorio Di Tomaso, Elisabeth Frasnelli, Ulrike Kugler, and B. Plank. 2006. Multilingual search in libraries. The case-study of the Free University of Bozen-Bolzano. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'06).* 2006, Genova, 2287-2290.

Marti Hearst, Ame Elliott, Jennifer English, Rashmi Sinha, Kirsten Swearingen, and Ka-Ping Yee. 2002. Finding the flow in web site search. *Communications of the ACM* 45, 9 (Sept. 2002), 42-49.

Gary Marchionini. 2006. Exploratory search: from finding to understanding. *Communications of the ACM - Supporting exploratory search* 49, 4 (April 2006), 41-46.

Roberto Navigli and Paola Velardi. 2003. An Analysis of Ontology-based Query Expansion Strategies. In *Proceedings of the 14th European Conference on Machine Learning, Workshop on Adaptive Text Extraction and Mining.* Cavtat-Dubrovnik, Croatia, 42-49.

Gerard Salton. 1984. The use of extended Boolean logic in information retrieval. In *Proceedings of the 1984 ACM SIGMOD international conference on Management of data (SIGMOD'84).* ACM, New York, NY, 277-285.

Osma Suominen, Kim Viljanen, and Eero Hyvänen. 2007. User-centric faceted search for semantic portals. *The Semantic Web: Research and Applications* 4519, 356-370.

Ryen W. White and Gary Marchionini. 2007. Examining the effectiveness of real-time query expansion. Information Processing & Management 43, 3 (May 2007), 685-704.

Questions