Universidade de São Paulo

TRIVIR: A Visualization System to Support Document Retrieval with High Recall

Student

Sherlon Almeida da Silva

sherlon@usp.br

Authors

Amanda G. Dias
gdias.amanda@gmail.com

Evangelos E. Milios
eem@cs.dal.ca

Maria C. F. de Oliveira
cristina@icmc.usp.br

Institute of Mathematical and Computer Sciences - ICMC/USP

Introduction

What is the Motivation?

Motivation

User

A scientist needs to discover relevant information about her research, e.g. studying the literature.

Motivation

User

Query

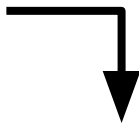She already knows some interesting papers that can be useful to start the search.

Motivation

She tries to find relevant information in a collection of papers (a corpus).
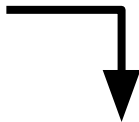
Query

User

Corpus

Motivation

She identifies relevant papers in the collection.

Query

User

Corpus

Relevant Docs

Introduction

What is the Problem?

Problem

Think about this situation

Problem

Think about this situation
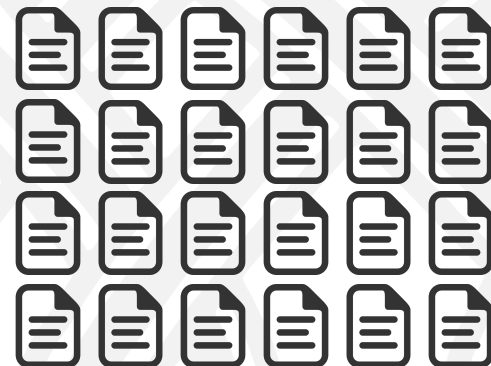
- There is a corpus containing lots of documents;

Problem

Think about this situation

- There is a Corpus containing lots of documents;
- Find similar or related (i.e., ´interesting´) documents;

Problem

Think about this situation

- There is a Corpus containing lots of documents;
- Find similar or related documents;
- How can you identify the relevant documents in a corpus with lots of documents?

Problem

Think about this situation

- There is a Corpus containing lots of documents;
- Find similar or related documents;
- How can you identify the **relevant** documents in a corpus with <u>**lots of documents**</u>?

Analyze ALL documents?

Problem

Think about this situation

- There is a Corpus containing lots of documents;
- Find similar or related documents;
- How can you identify the relevant documents in a corpus with lots of documents?

Machine Learning (ML)
+ Visual Analytics (VA)

TRIVIR

The Visualization System
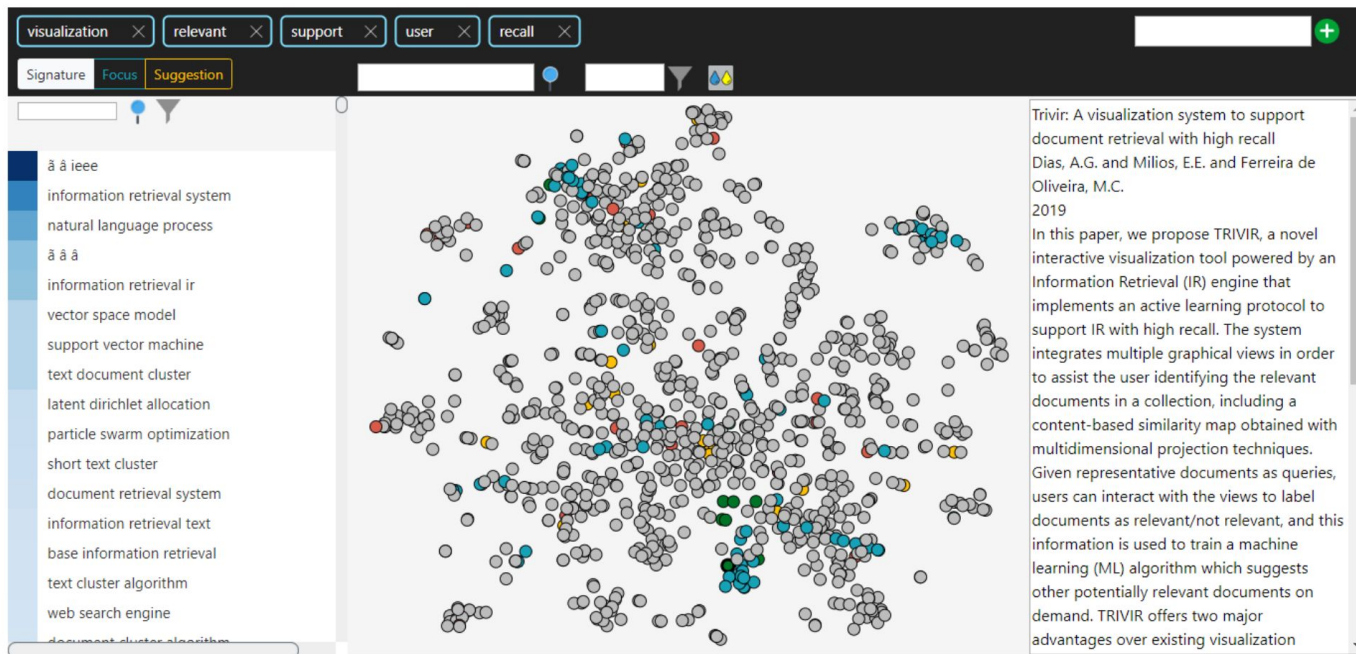
TRIVIR

Visualization System
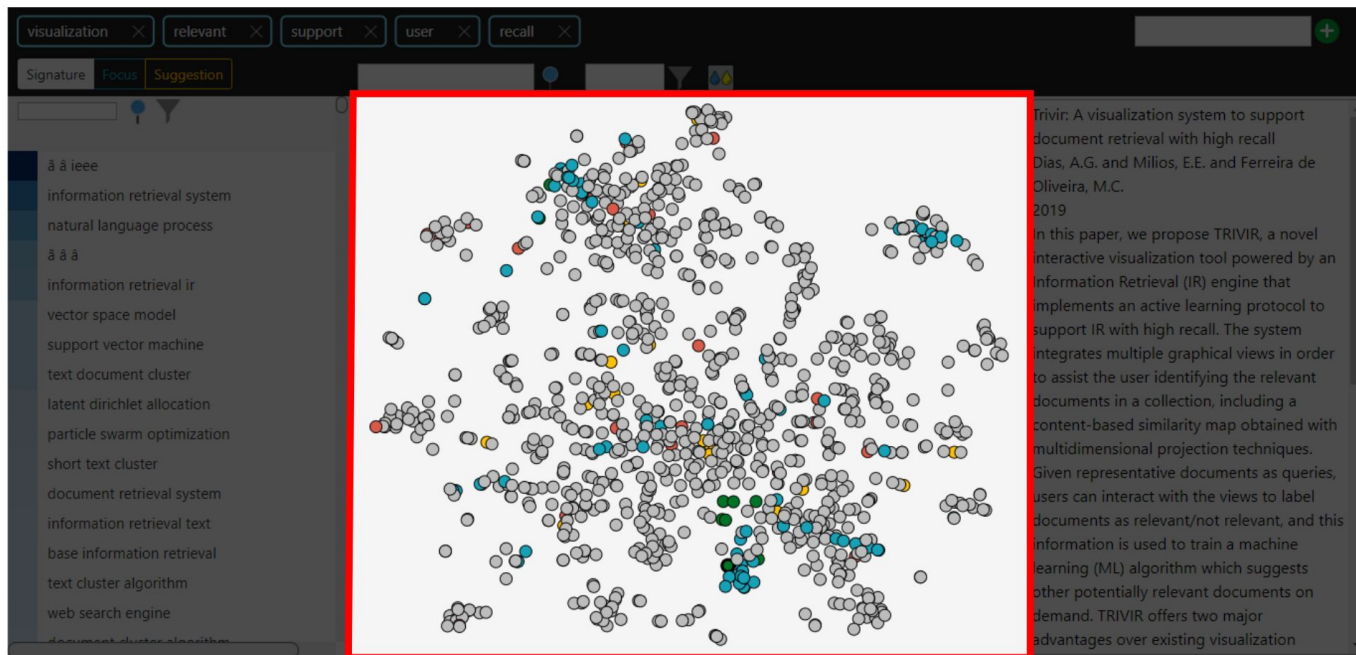
TRIVIR

Scatterplot view:
Similarity Map of Documents

TRIVIR
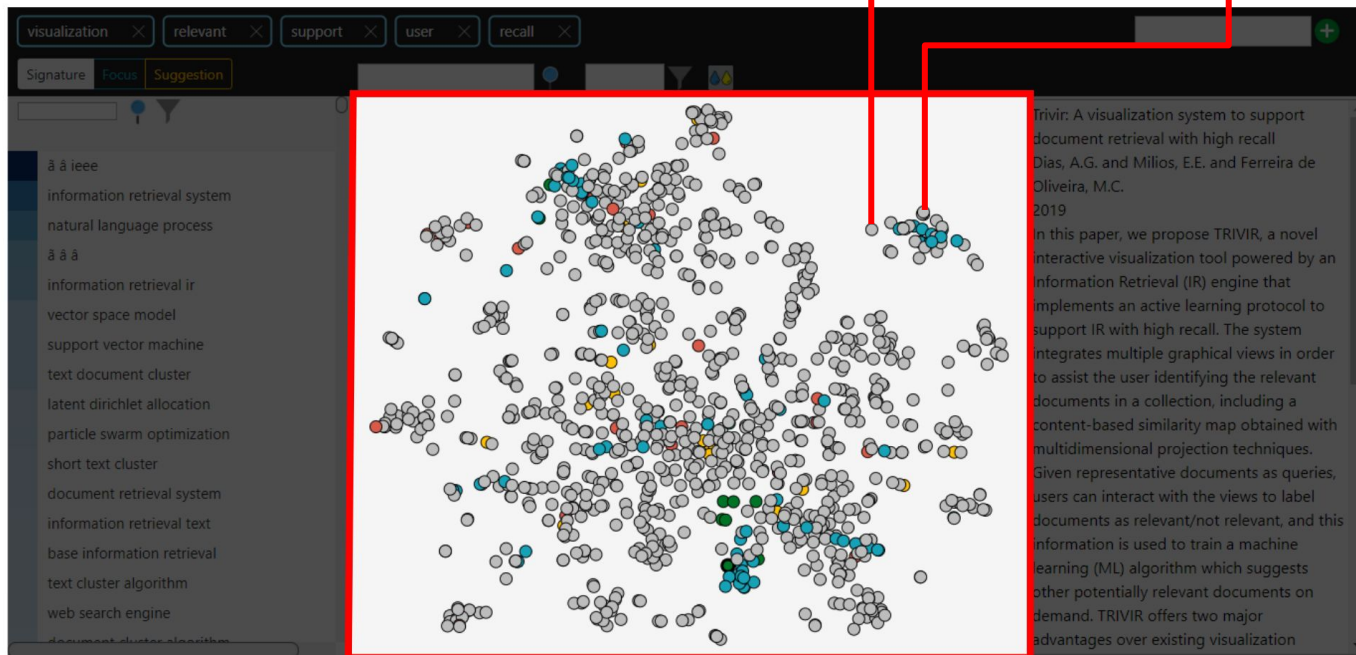
Documents

Scatterplot view:
Similarity Map of Documents

TRIVIR

Clusters of Related Documents
Multidimensional Projection

Scatterplot view:
Similarity Map of Documents

18
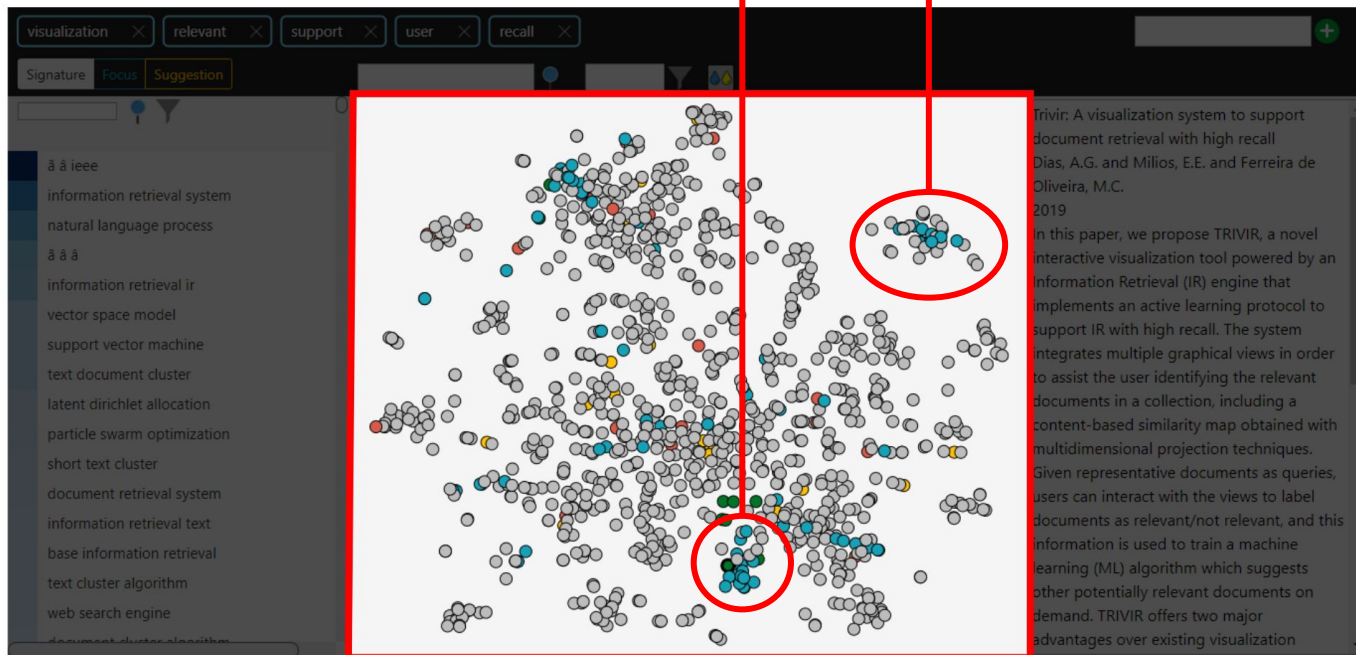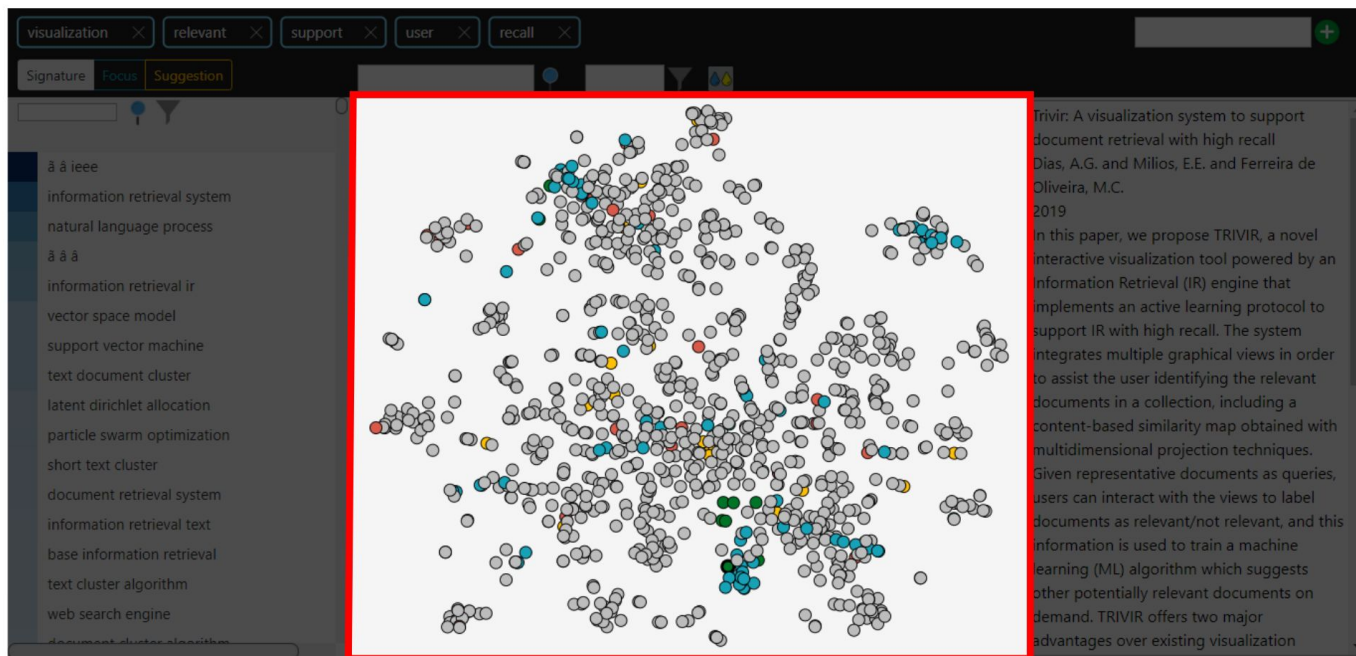
TRIVIR

- 🟢 Current Query
- 🔵 Relevant
- 🔴 Not Relevant
- 🟡 Suggested
- ⚪ Unlabeled Documents

TRIVIR
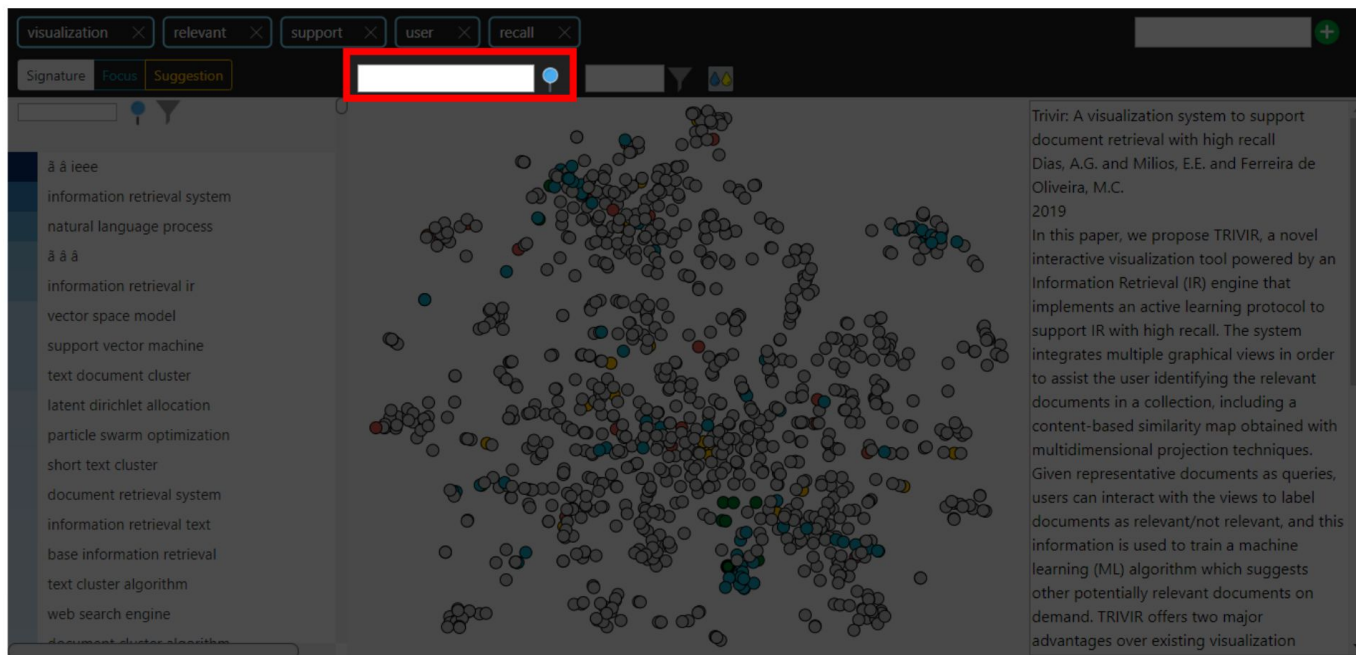
Select Documents that include a given term

TRIVIR
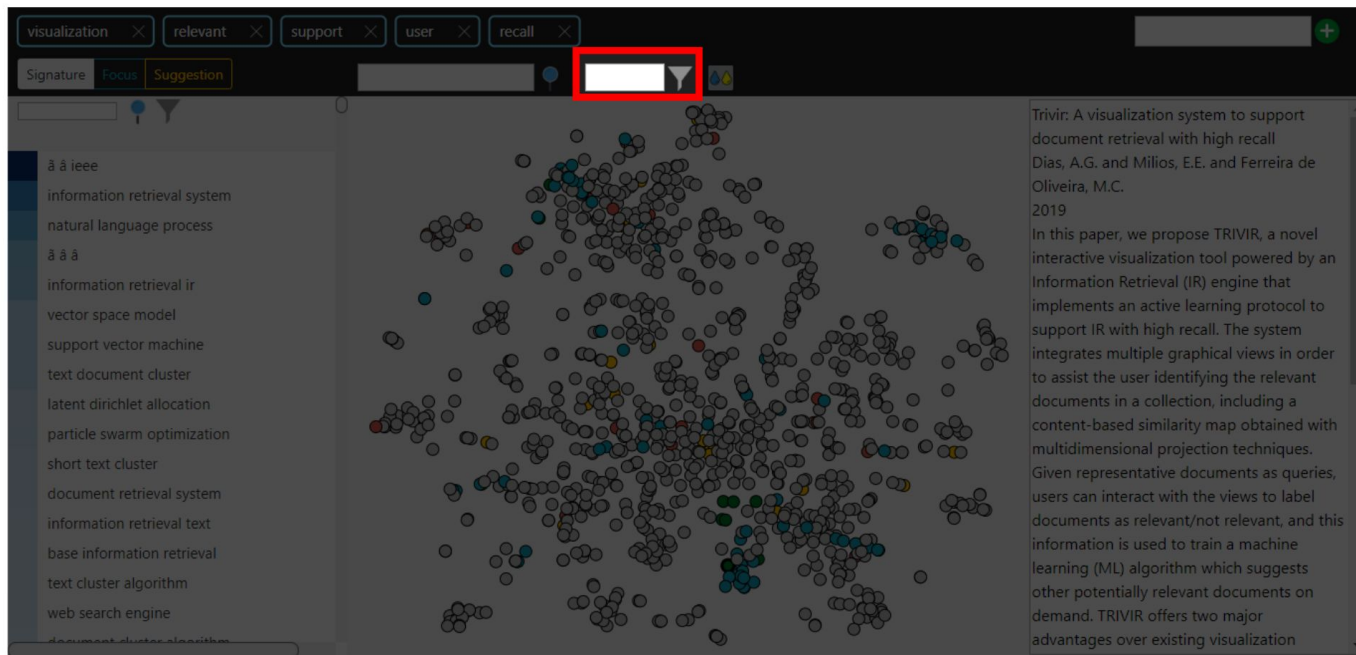
Filter the K documents <u>most similar</u> to the query

TRIVIR
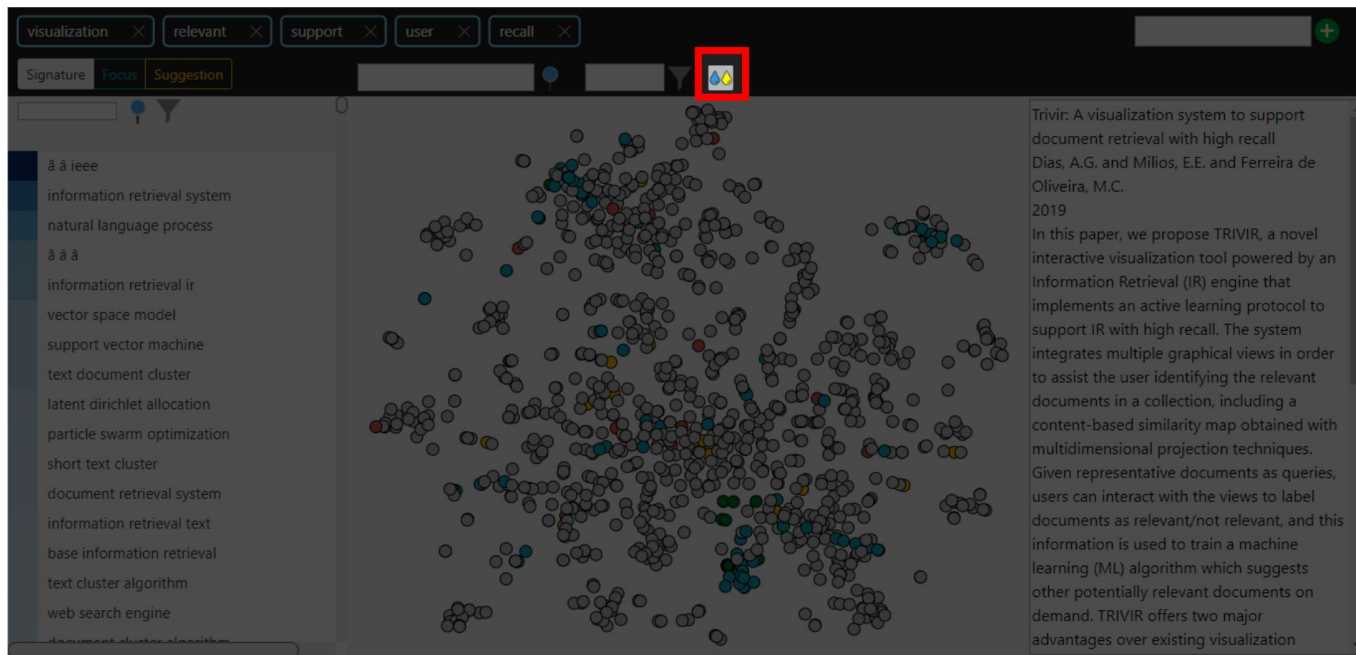
Reduce point clutter by displaying only the
seed, relevant and suggested documents

TRIVIR
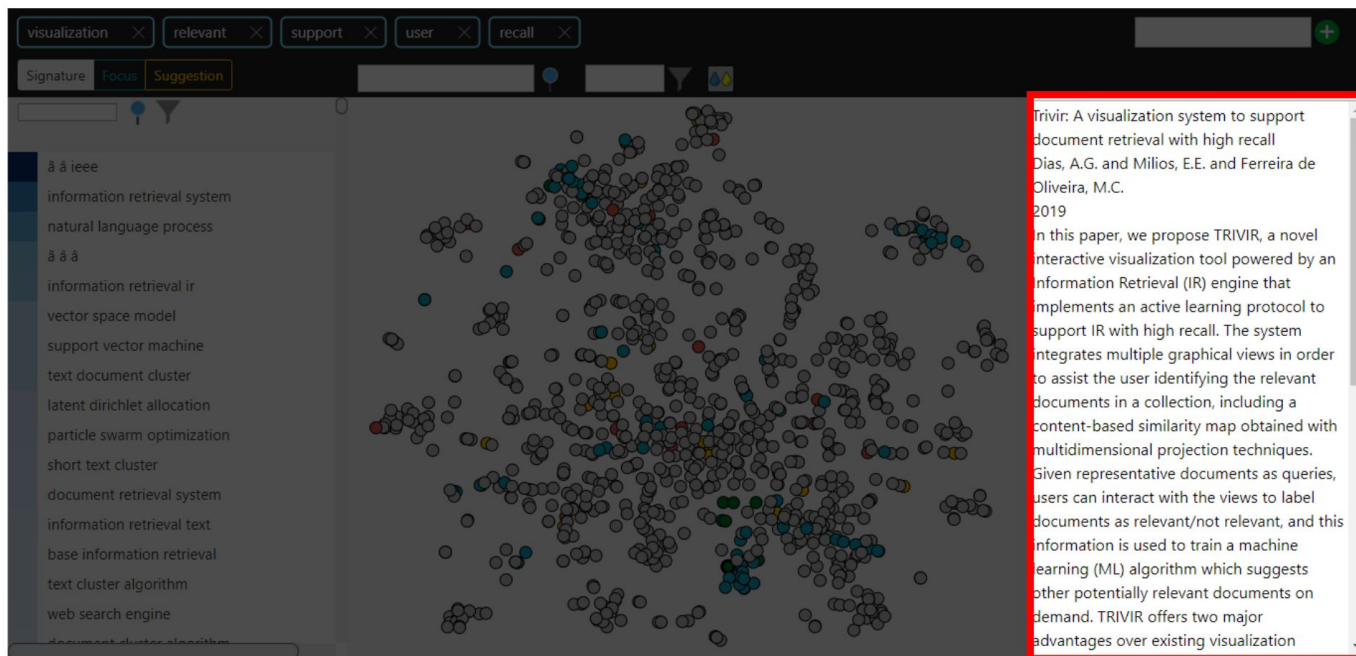
Query document shown:
Clicking on a circle, the document's content is displayed in the Document view

TRIVIR
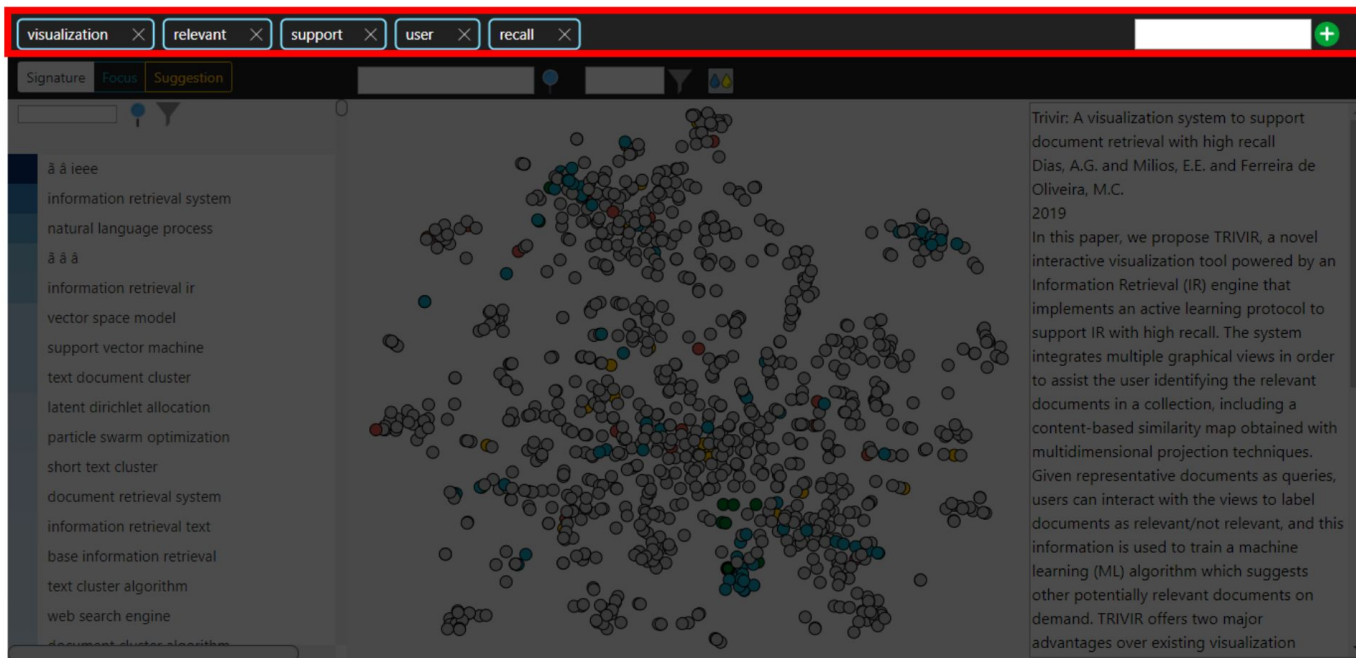
The Terms view:
Shows important terms from the query document,
where the user can remove or add terms.

TRIVIR
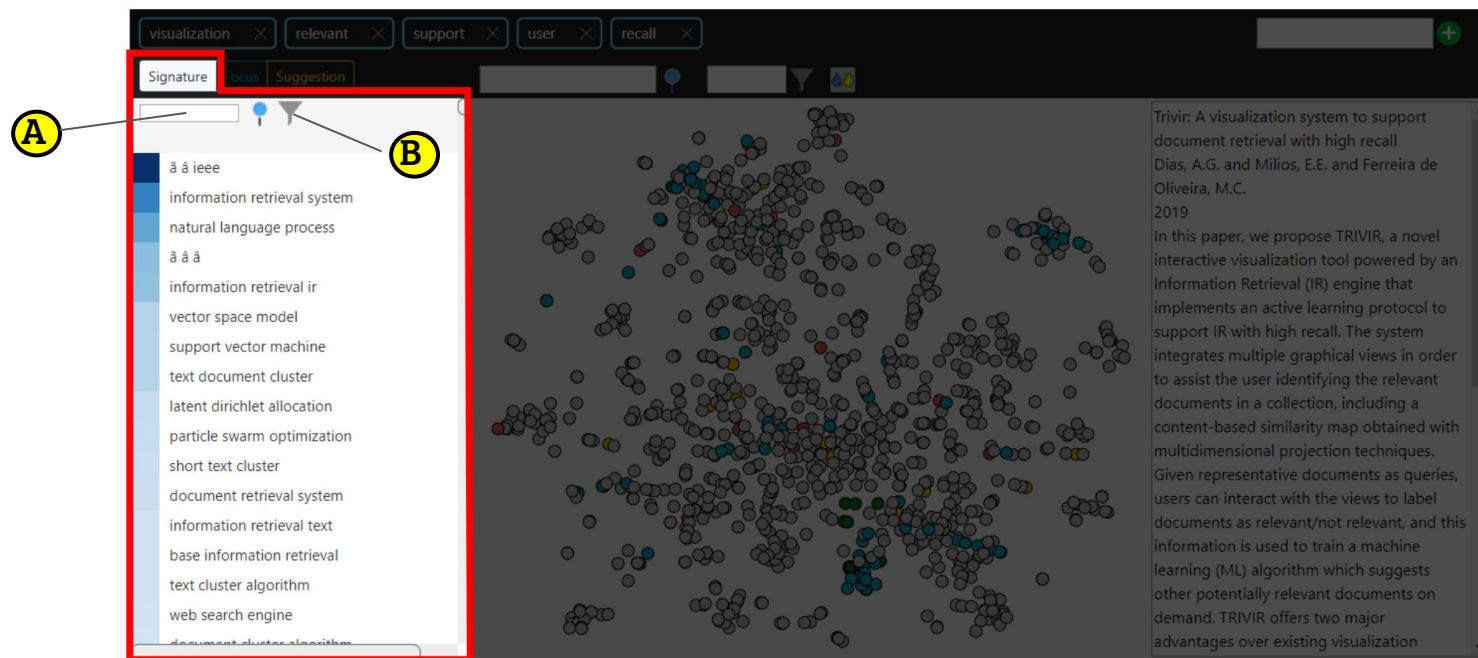
Signature List view:
Shows a rank of <u>relevant 3-grams</u> in the corpus. The user can select 3-grams
with a <u>particular term</u> (A) or with <u>terms from the Terms view</u> (B) .

TRIVIR

Focus List view:
Shows all documents currently labeled as seed and relevant.

TRIVIR

Suggestion List view:
Shows the documents suggested as relevant
by the ML classifier.

TRIVIR

Suggestion List view:

A Retrain the ML algorithm considering your previous feedback. This step can improve the recommendation of documents.

User Study

Hands On!

User Study

Thinking aloud:

- **Describe your thoughts**

Make appointments:

- Strengths
- Weaknesses
- General Considerations