



Universidade de São Paulo

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

Student

Sherlon Almeida da Silva

sherlon@usp.br

Amanda G. Dias

gdias.amanda@gmail.com

Evangelos E. Milios

eem@cs.dal.ca

Authors

Maria C. F. de Oliveira

cristina@icmc.usp.br

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.

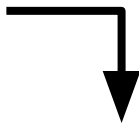
Motivation



User



Query



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



Corpus

Motivation

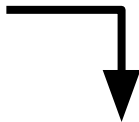


User



Query

She identifies relevant papers in the collection.



Corpus



Relevant Docs

Introduction

What is the Problem?

Problem

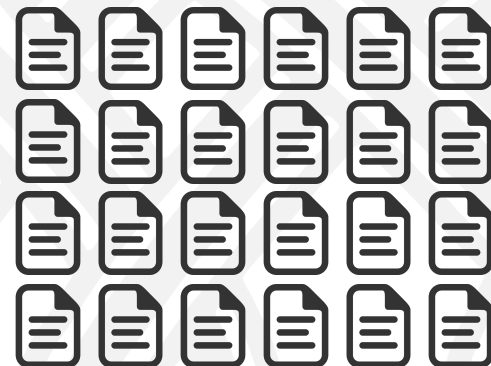
Think about this situation



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 **lots of documents**?



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?**

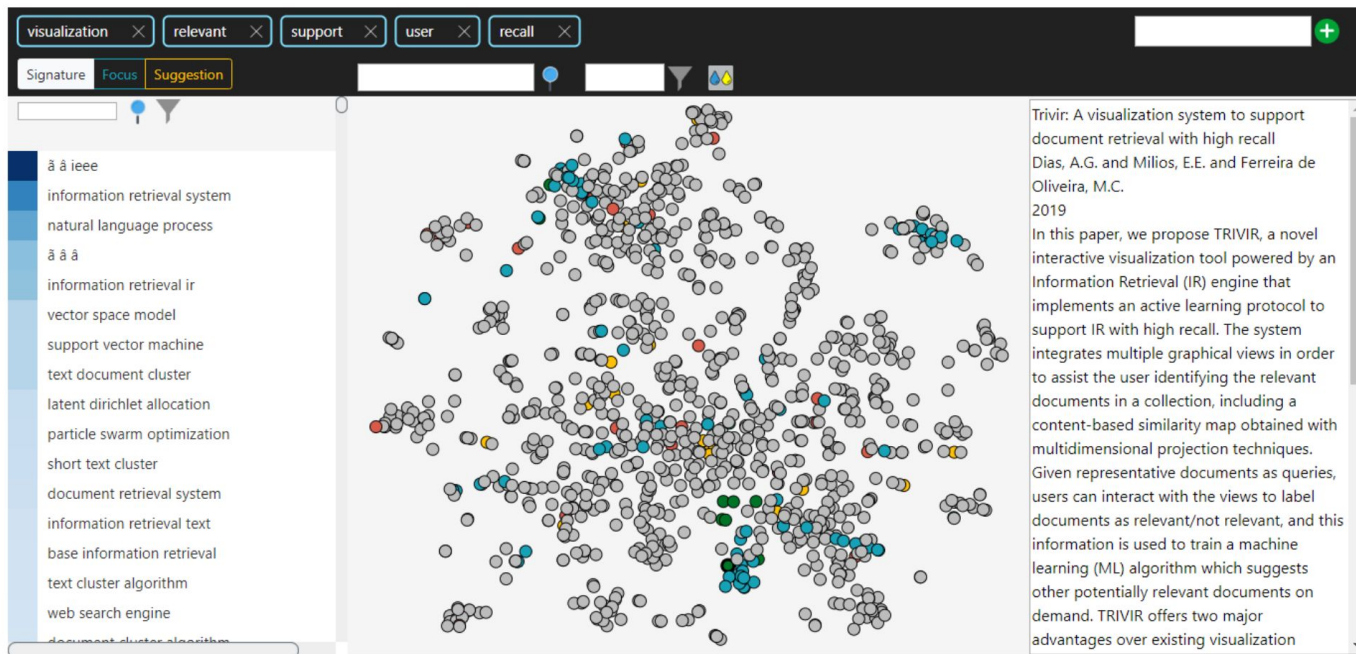


Machine Learning (ML)
+ Visual Analytics (VA)

TRIVIR

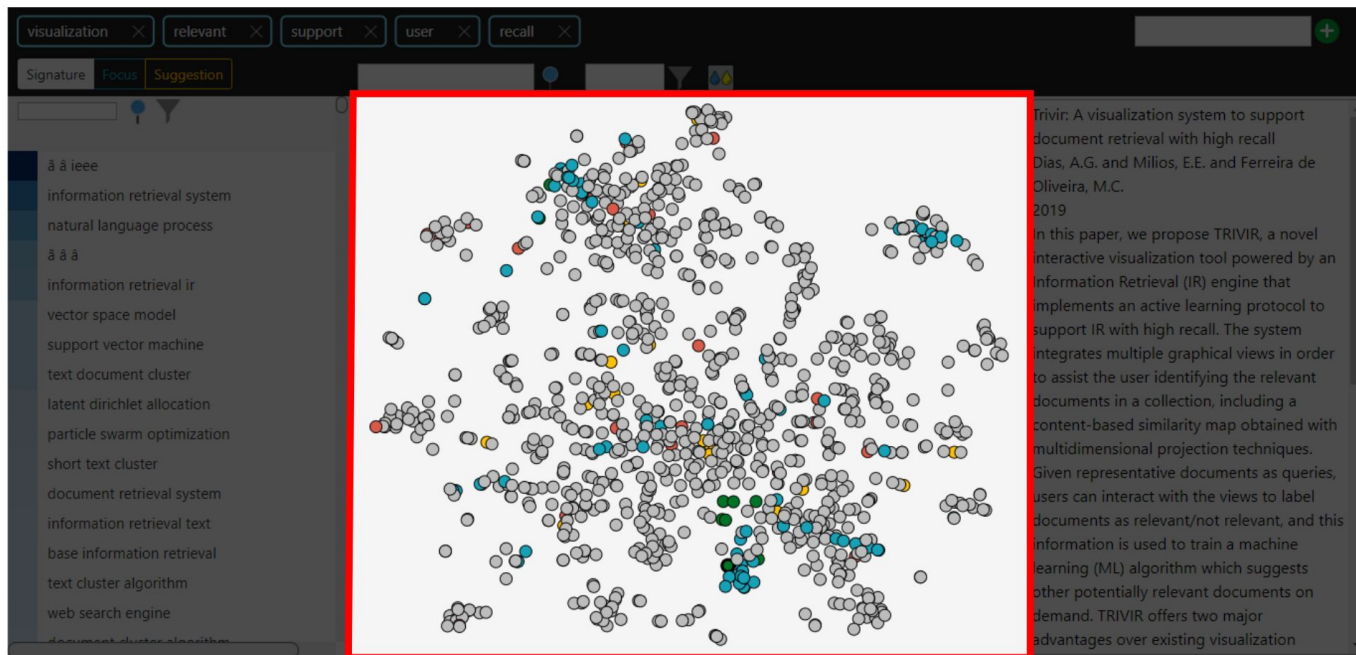
The Visualization System

TRIVIR



Visualization System

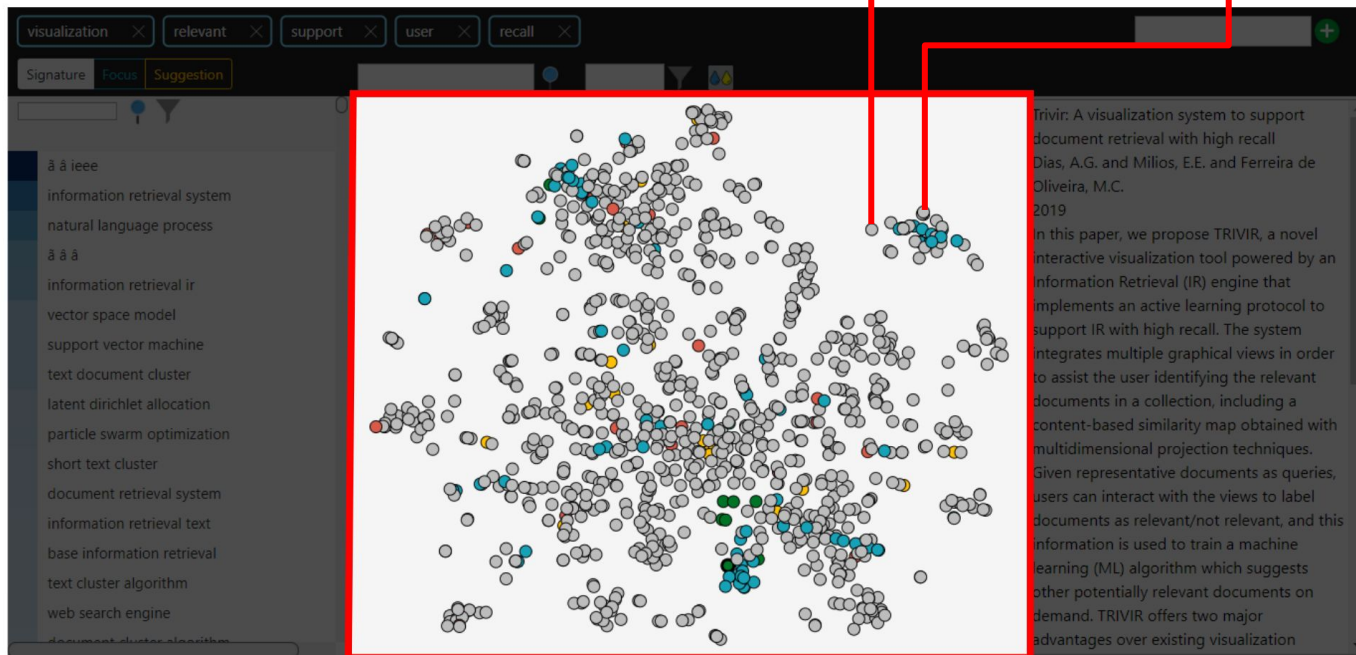
TRIVIR



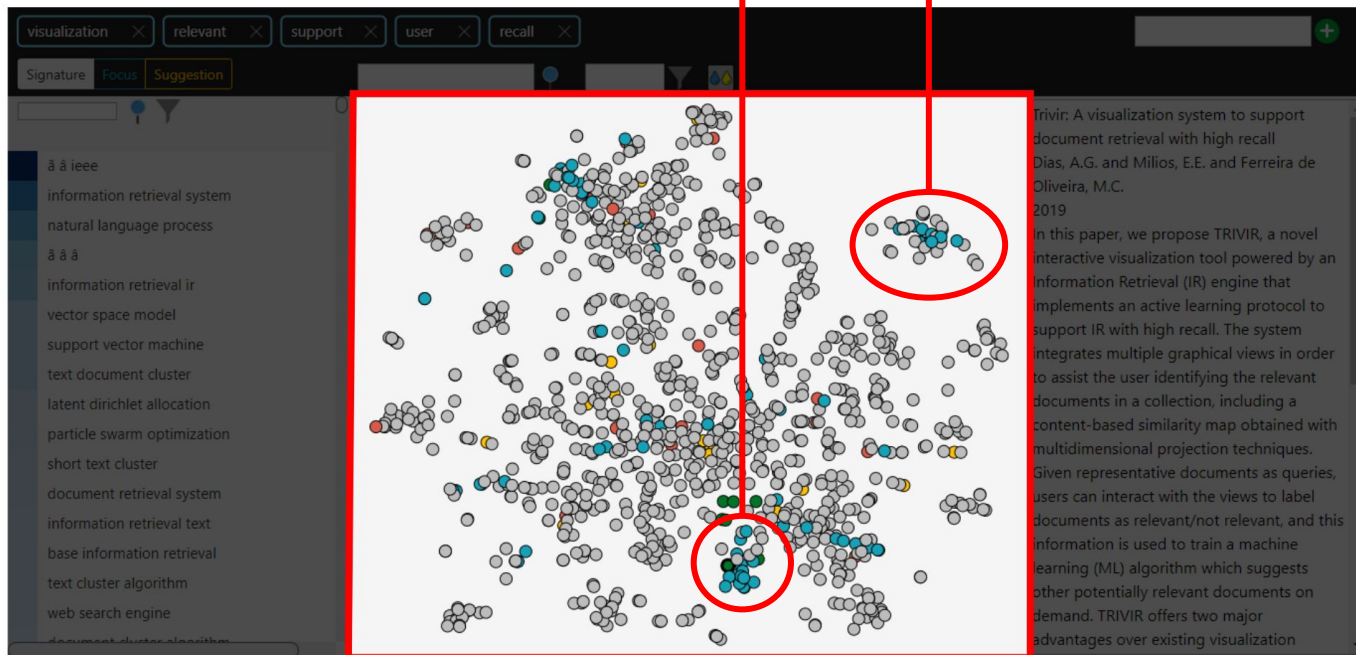
Scatterplot view:
Similarity Map of Documents

TRIVIR

Documents

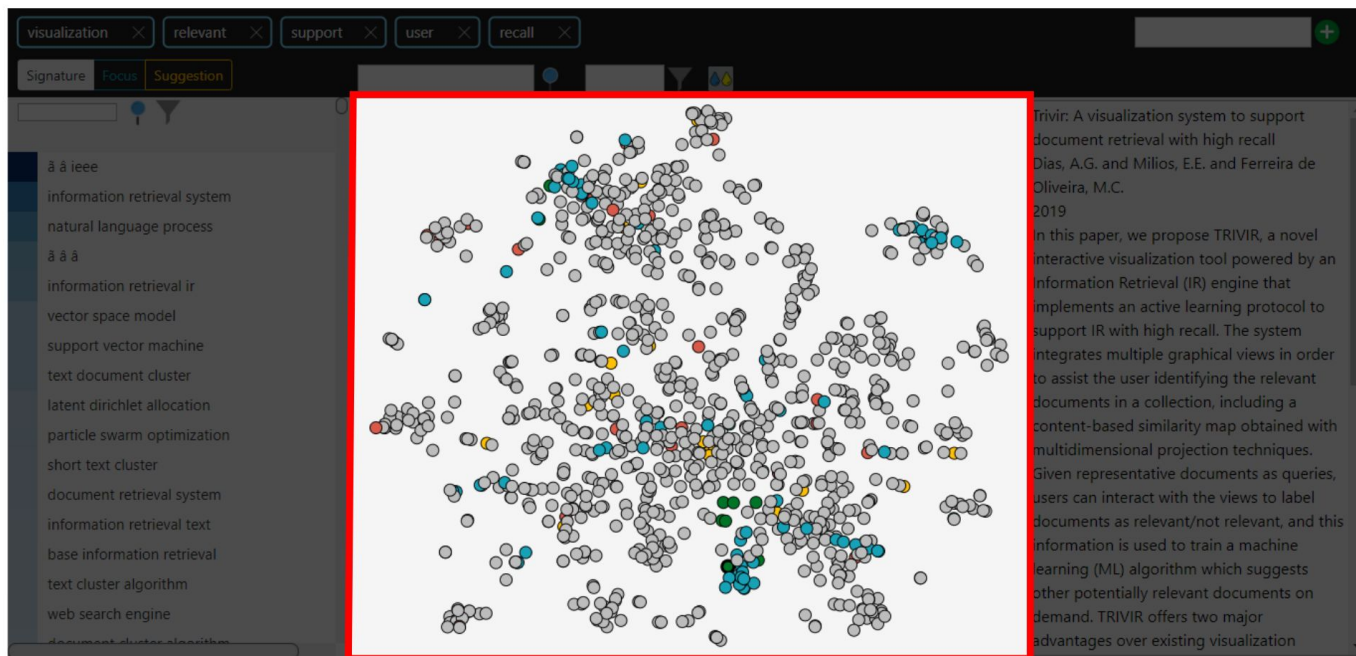


Scatterplot view:
Similarity Map of Documents



Scatterplot view:
Similarity Map of Documents

TRIVIR



● Current Query

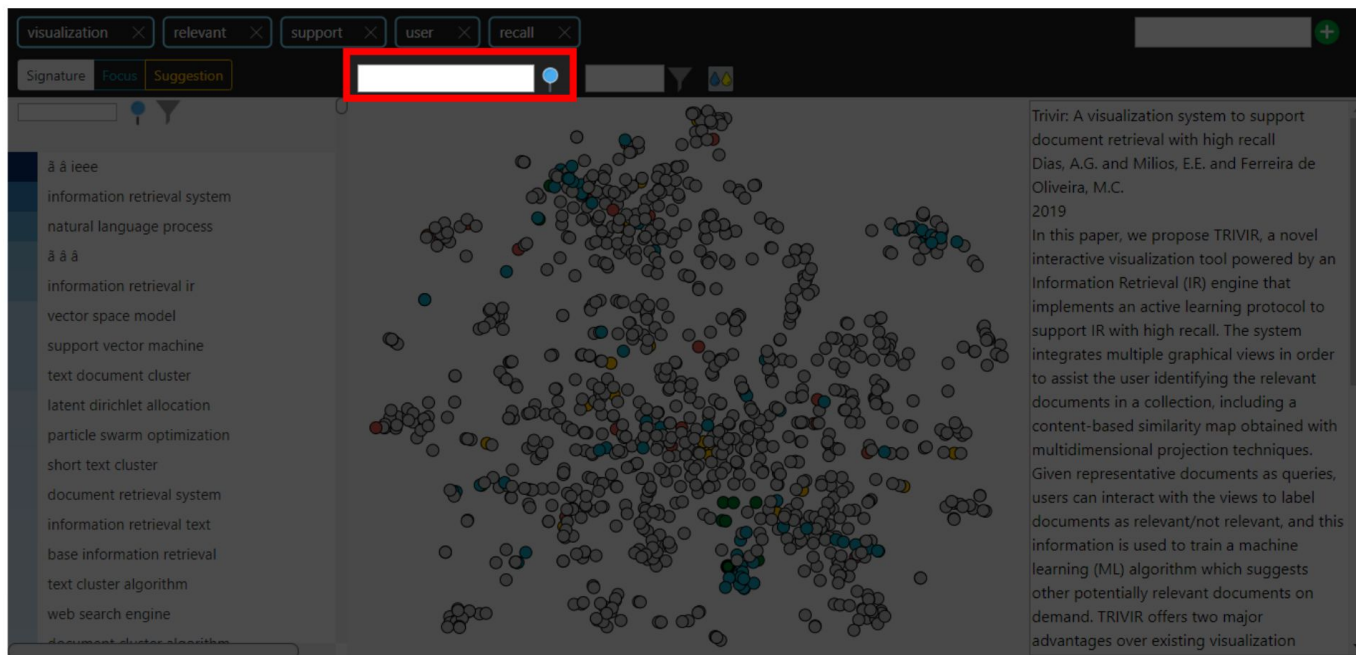
● Relevant

● Not Relevant

● Suggested

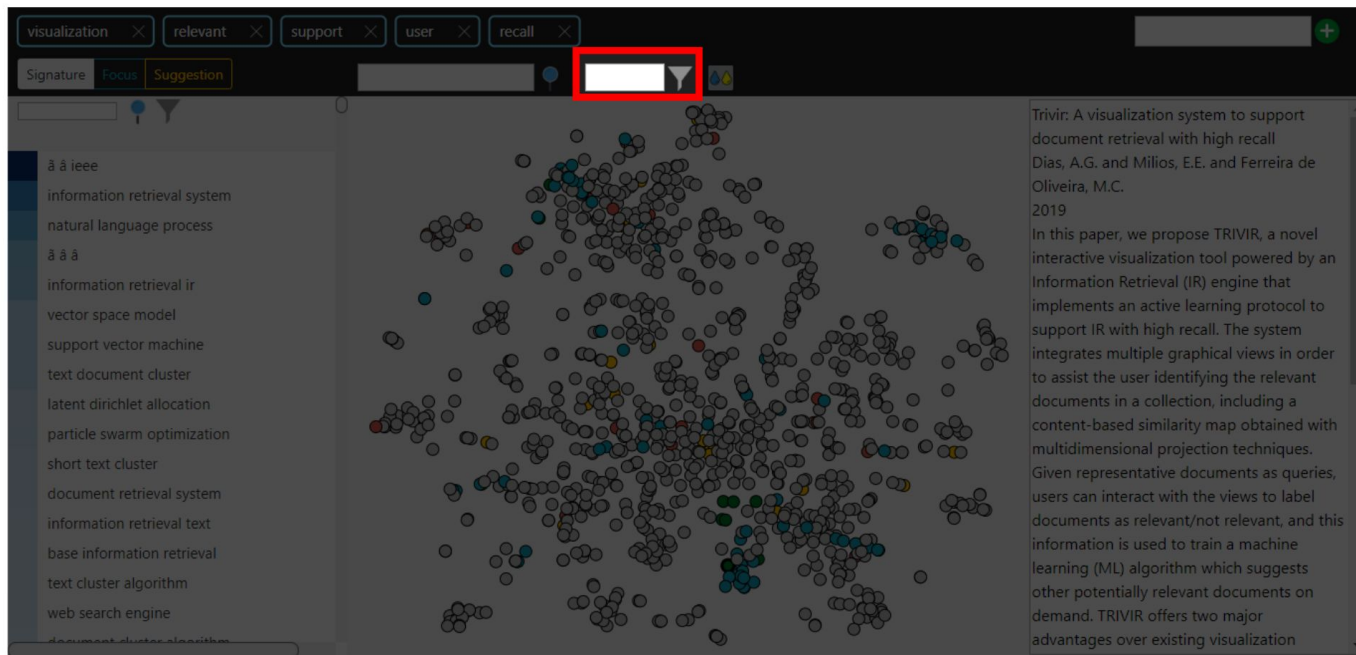
● Unlabeled Documents

TRIVIR



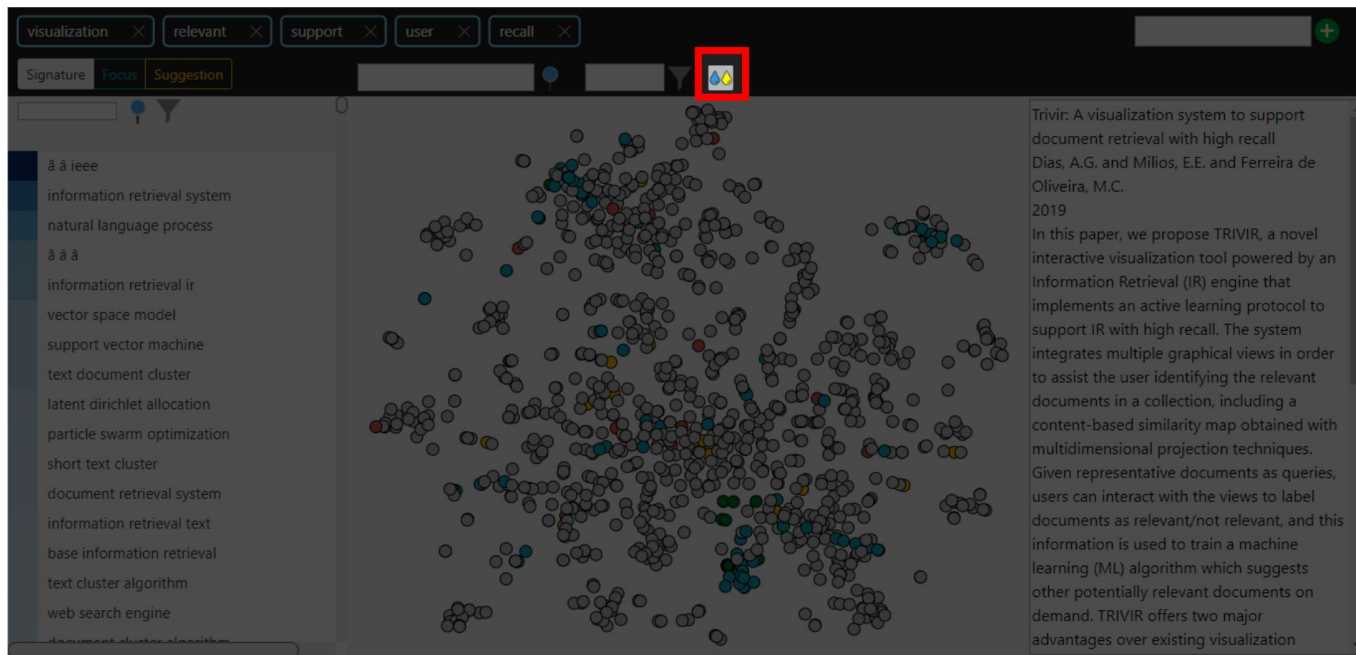
Select Documents that include a **given term**

TRIVIR



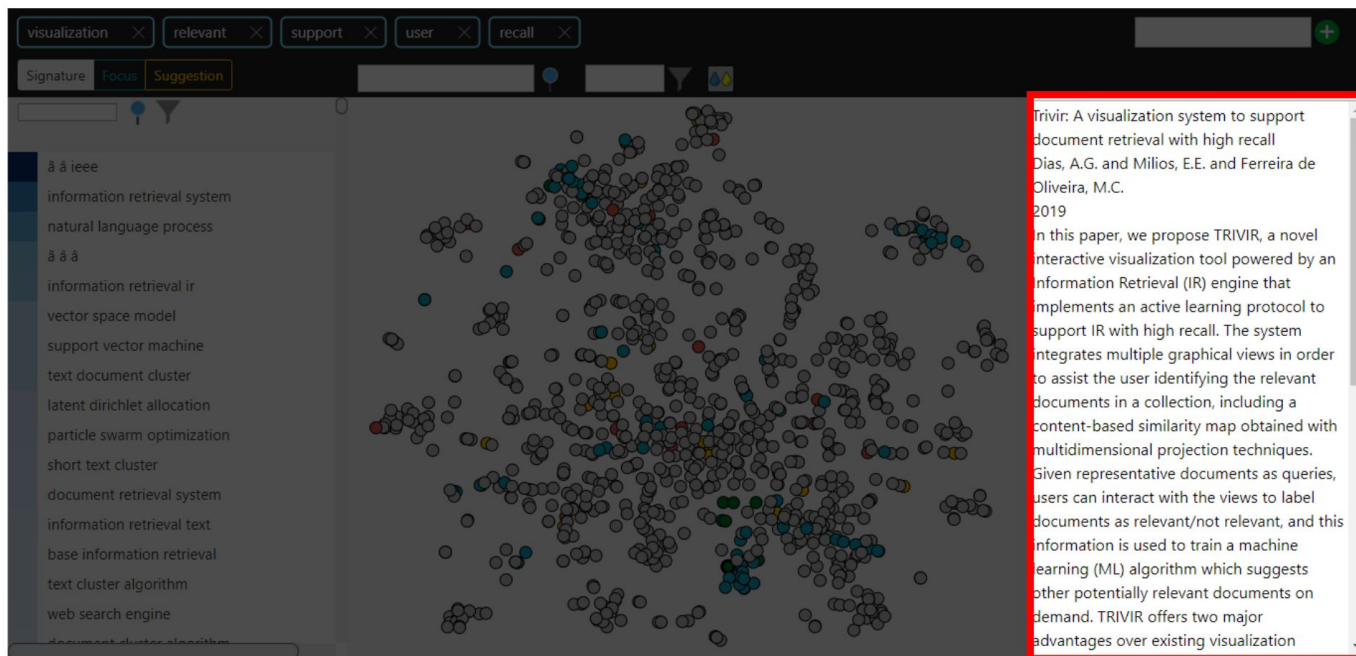
Filter the **K** documents most similar to the query

TRIVIR



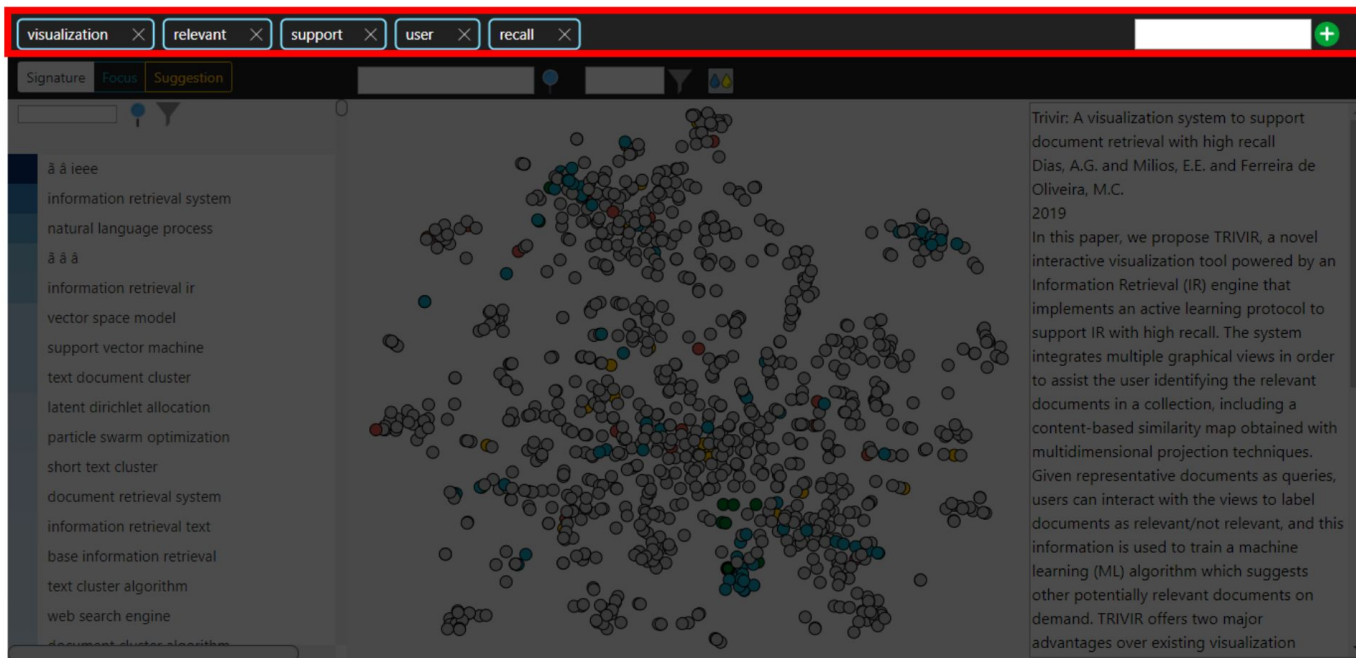
Reduce point clutter by displaying only the **seed, relevant** and **suggested** documents

TRIVIR



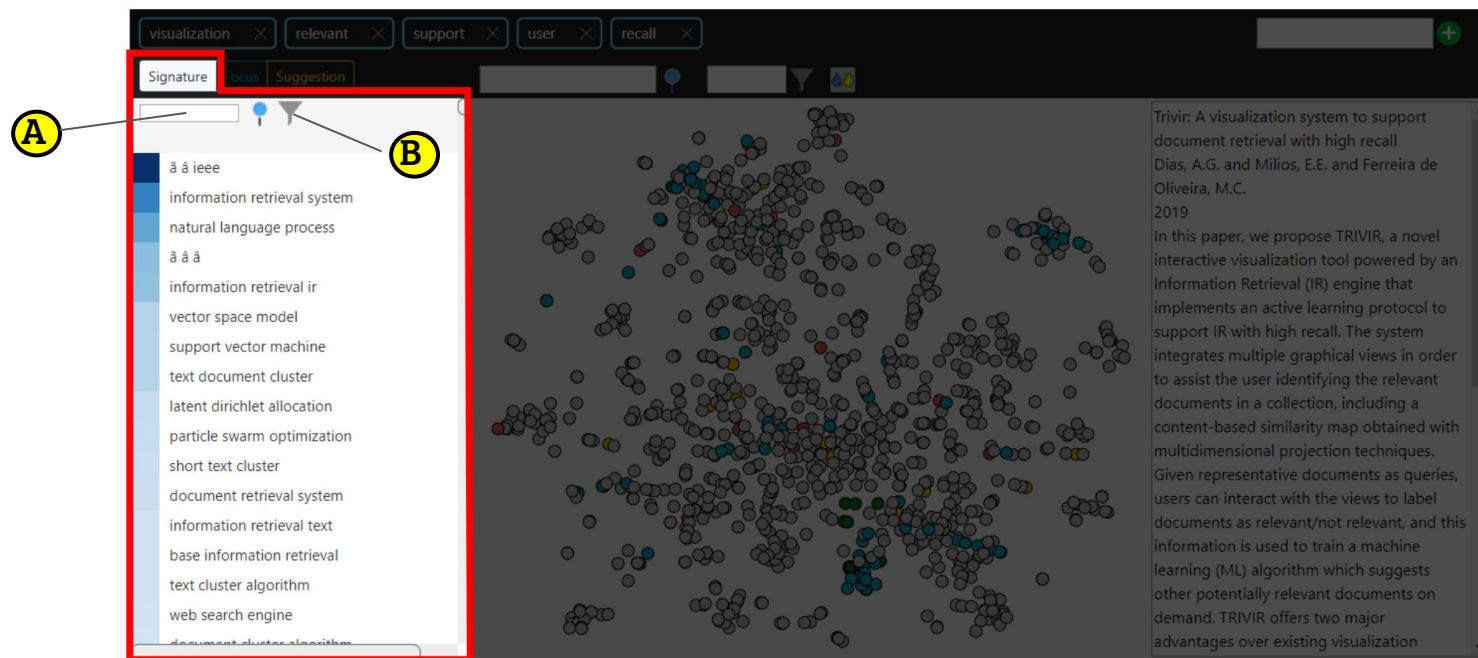
Query document shown:
Clicking on a circle, the document's content is displayed in the Document view

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 relevant 3-grams in the corpus. The user can select 3-grams with a particular term **A** or with terms from the Terms view **B**.

TRIVIR

visualization × relevant × support × user × recall ×

Signature **Focus** Suggestion

total: 86

- Trivir: A visualization system to support document retrieval with high recall
- Trivir: A visualization system to support document retrieval with high recall
- Interpretive Impacts of Text Visualization: Mitigating Cognitive Load
- ViStA: a visualization system for exploring Arabic text
- Digging Text Viz: An Archaeological Review of Ancient Texts
- Sos textvis: An extended survey of surveys on text visualization
- Doccurate: A Curation-Based Approach for Clinical Document Clustering
- A New Evolving Tree-Based Model with Local Refinement
- Strategies for evaluating information visualization
- An Integrated Visualization System for Spatial Data
- Seeing beyond reading: A survey on visual text analysis
- TopicVis: Visualizing complex topic models for user exploration
- Interactive document clustering revisited: A visualization-based approach
- GF-CLUST: A nature-inspired algorithm for automatic document clustering
- Documents clustering using quantum clustering
- Document clustering using gravitational ensemble learning
- Scalability analysis of semantics based distributed document clustering

Trivir: A visualization system to support document retrieval with high recall
Dias, A.G. and Milios, E.E. and Ferreira de Oliveira, M.C.
2019
In this paper, we propose TRIVIR, a novel interactive visualization tool powered by an Information Retrieval (IR) engine that implements an active learning protocol to support IR with high recall. The system integrates multiple graphical views in order to assist the user identifying the relevant documents in a collection, including a content-based similarity map obtained with multidimensional projection techniques. Given representative documents as queries, users can interact with the views to label documents as relevant/not relevant, and this information is used to train a machine learning (ML) algorithm which suggests other potentially relevant documents on demand. TRIVIR offers two major advantages over existing visualization

Focus List view:

Shows all documents currently labeled as **seed** and **relevant**.

TRIVIR

visualization × relevant × support × user × recall ×

Signature Focus **Suggestion**

Train

- Monolingual Information Retrieval using Terrier:
- Methods and trends of biomedical and genomic
- ConceptVector: Text Visual Analytics via Interactiv
- User Profiling for University Recommender Syste
- Map Archive Mining: Visual-Analytical Approach
- Link based BPSO for feature selection in big data
- Arabic Information Retrieval Using Semantic Ana
- Machine transliteration and transliterated text re
- Information retrieval in physical geography: A m
- A framework for intelligent question answering s
- Joint embeddings with multimodal cues for vide
- Let's get personal: the little nudge that improves
- Lempel-Ziv compressed structures for document
- Preceding Document Clustering by Graph Mining
- Estimation of genetic diversity and relatedness in
- A hybrid approach using genetic algorithm and t
- Using bug descriptions to reformulate queries du

Trivir: A visualization system to support document retrieval with high recall
Dias, A.G. and Milios, E.E. and Ferreira de Oliveira, M.C.
2019
In this paper, we propose TRIVIR, a novel interactive visualization tool powered by an Information Retrieval (IR) engine that implements an active learning protocol to support IR with high recall. The system integrates multiple graphical views in order to assist the user identifying the relevant documents in a collection, including a content-based similarity map obtained with multidimensional projection techniques. Given representative documents as queries, users can interact with the views to label documents as relevant/not relevant, and this information is used to train a machine learning (ML) algorithm which suggests other potentially relevant documents on demand. TRIVIR offers two major advantages over existing visualization

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

TRIVIR

visualization × relevant × support × user × recall ×

Signature Focus **Suggestion**

Train

- Monolingual Information Retrieval using Terrier:
- Methods and trends of biomedical and genomic
- ConceptVector: Text Visual Analytics via Interactiv
- User Profiling for University Recommender Syste
- Map Archive Mining: Visual-Analytical Approach
- Link based BPSO for feature selection in big data
- Arabic Information Retrieval Using Semantic Ana
- Machine transliteration and transliterated text re
- Information retrieval in physical geography: A m
- A framework for intelligent question answering s
- Joint embeddings with multimodal cues for vide
- Let's get personal: the little nudge that improves
- Lempel-Ziv compressed structures for document
- Preceding Document Clustering by Graph Mining
- Estimation of genetic diversity and relatedness in
- A hybrid approach using genetic algorithm and t
- Using bug descriptions to reformulate queries du

Trivir: A visualization system to support document retrieval with high recall
Dias, A.G. and Milios, E.E. and Ferreira de Oliveira, M.C.
2019
In this paper, we propose TRIVIR, a novel interactive visualization tool powered by an Information Retrieval (IR) engine that implements an active learning protocol to support IR with high recall. The system integrates multiple graphical views in order to assist the user identifying the relevant documents in a collection, including a content-based similarity map obtained with multidimensional projection techniques. Given representative documents as queries, users can interact with the views to label documents as relevant/not relevant, and this information is used to train a machine learning (ML) algorithm which suggests other potentially relevant documents on demand. TRIVIR offers two major advantages over existing visualization

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