NGI Search: The need for Trust and Privacy in Search, Discovery and Indexing

Aurora González-Vidal[1], Antonio F. Skarmeta[1], Mirko Presser[2], Marie Claire Tonna[3], Manuel Noya[4], Pierre-Yves Gibello [5]

[1] DIIC, University of Murcia, Spain
[2] DBDT, Aarhus University, Denmark
[3] FundingBox Accelerator, Poland
[4] Linknovate Science S.L., Spain
[5] OW2, France

October 11, 2022

Overview

Introduction and motivation
○●○○

Workprogramme
○○○○○○○○○○○○○○

Methodology
○○○○

Conclusion
○○

Introduction and motivation

What is NGI Search?

The NGI Search project is a European project designed to support entrepreneurs, tech-geeks, developers, and socially engaged people, who are capable of challenging the way we **search and discover** information and resources on the internet.

Motivation

- Applications using internet and connected devices →search and discovery a challenge
- Find, represent and make-sense of resources in a transparent and trustworhty way
- Mission: to help develop technologies and solutions enabling new and trustworthy ways of searching and discovering information on the internet

Introduction and motivation
○○○●

Workprogramme
○○○○○○○○○○○○○○

Methodology
○○○○

Conclusion
○○

Open Calss

NGI Search will organise five Open Calls for innovative projects whereby selected applicants will benefit from the *financial support* of up to 150,000€ as well as *technical, business and innovation support* over 12 months.

Introduction and motivation
OOOO

Workprogramme
●OOOOOOOOOOOOO

Methodology
OOOO

Conclusion
OO

Workprogramme

Topics

1st open call topics

- Intelligent Voice Based Assistants
- Natural Language Processing
- Semantic analysis
- Social computing
- Data visualization
- New ways of discovering and accessing information

Other topics

- Power cognitive search by reinforcement learning
- Machine-based data (IoT)
- Network analysis - knowledge graphs
- AI-based taxonomies
- Addressing verticals
- Other topics
  - Federated Search
  - Transfer Learning
  - Data segmentation and representation methods
  - Information retrieval

Intelligent Voice Based Assistants

Voice assistants are used for routine tasks and as familiarity
increases, more complex tasks will require the addition of
intelligence to the assistants.

- Make them context-aware
- Mixed modal interactions: input and output are not only voice but also images or sounds
- Privacy concerns: when searches are done in public, the response should keep ones privacy
- Interpret tone
- Multilangual queries

Natural Language Processing

(NLP) methods are widely in use. The majority of search engines use the huge amounts of previously accumulated user requests for predicting the search output Model complexity of the current models is increasing and implies the use of great amounts of energy for computation.

Simpler and more sustainable NLP research and practices

- Edge/Tiny computing for NLP
- Federated NLP
- Social Media data

Semantic analysis

Semantic data integration generates a common representation of concepts and their relations, aiding in the integration of heterogeneous data. Information about a subject or topic might be spread across different data sources so there exists the need for the integration of the knowledge.

- Search dynamic relations between concepts
- Federated query processing for large-scale heterogeneous data

Social computing

Social computing are the systems that support the gathering, representation, processing, use, and dissemination of information that is distributed across social collectivities.

The interaction between social behaviour and technologies needs to be addressed in order to reach substantial changes in the behaviour of the adopters.

Edge social computing, collaborative learning and other proposals in this direction are encouraged.

Data visualization

Data visualization has attracted much attention recently, calling
for joint actions in different research fields such as information
visualization, human-computer interaction, machine learning, data
management and mining, and computer graphics. We seek
interactive tools and mechanisms that allow visualizations for
machine learning / query results that can provide user
recommendations and support user-driven actions (all kinds of
scalable and efficient solutions)

- Visually-driven analysis of spatio-temporal, textual and other
  kinds of data
- progressive visualizations (in batches)

New ways of discovering and accessing information

Due to the rapid development of the IoT and the variability and volume of data sources, mechanisms for searching and integrating data are essential to leverage all relevant knowledge for improving processes and services.

The integration of data-driven machine learning with human knowledge can effectively lead to explainable AI that would provide us ways to discover and access information where only raw data is present.

The Challenge is to develop new algorithms and methodologies to discover and access information by combining Big Data technologies.

Power cognitive search by reinforcement learning

Development of mechanisms, including but not restricting to self learning, pattern recognition and natural language processing, that contribute to a reinforcement learning system able to learn from the interactions how to choose the data and algorithms to make a search more relevant.

Machine-based data (IoT)

Enable the search and discovery of information based on historical data and pattern extraction by means of algorithms that can adapt to the characteristics of the IoT sources: geospatial information, events and time series, real-time, social media data, etc.

AI-based taxonomies

Taxonomies consist of machine-interpretable semantics and provide valuable knowledge for many applications such as enhancing query understanding.

The existing generic taxonomies cannot satisfy user's specific interests.

We encourage research about the automatic creation and expansion of taxonomies by means of AI techniques that model inter-dependency among new concepts.

Network analysis - knowledge graphs

The output of complex network analysis is an interlinked network
of distributed resources which can be queried, structures known as
knowledge graphs - very related to semantic modelling.
Formal semantics to derive conclusions in a query should be
addressed. Scalability, quality of the induced models, diversity on
the managed data and dynamicity are also general challenges of
knowledge graphs to account for.

Introduction and motivation
oooo

Workprogramme
oooooooooooooo●o

Methodology
oooo

Conclusion
oo

Addressing verticals

The challenge consists of proposing solutions to discover and compose services implemented across verticals.

- Smart cities
- Smart buildings / homes
- Industrial IoT
- Transportation & logistics
- Smart agriculture
- e-health

Other topics

- Federated Search, enabling a user to search several different data sources at once by making a single query
- Transfer Learning, that transfers the knowledge obtained using AI between domains with different characteristics
- Data segmentation and representation methods, that reduce the dimensionality of data while maintaining the information that it contains, easing search and discovery
- Information retrieval: Conversational search, that refers to the use of complete sentences and verbal units in search queries, zero-query search, that are systems that push information to the users based on their context and not on a specific query, and reproducibility of search

Introduction and motivation
oooo

Workprogramme
ooooooooooooo

Methodology
●ooo

Conclusion
oo

Methodology

Technical support

- Technology mentoring and advice on technologies
- Beta testing — the project will leverage ReachOut beta-testing platform[1] to support NGI-Search beneficiaries
- Efforts to link to Standards and Foundations

---

[1]https://www.reachout-project.eu/view/Main/

Business support

- Market Readiness Level[2] provides a Market Readiness Programme that facilitates adoption of open source by mainstream decision-makers.
- Pitch training
- Business modeling and coaching

---

[2]https://www.ow2.org/view/MRL/

Innovation management

- Open source licensing
- Market landscaping and research
- Open science advice

Introduction and motivation
0000

Workprogramme
00000000000000

Methodology
0000

Conclusion
●○

Conclusion

Introduction and motivation
oooo

Workprogramme
oooooooooooooo

Methodology
oooo

Conclusion
o●

Conclusion

For the next 3 years, 5 open calls will offer the opportunity to perform equity-free research and development on search, indexing and discovery. NGI Search projects will foster strong synergies with the NGI mission on a more human centric internet with focus on privacy and trust as key concepts.
The core values are

- open source
- contributions to the wider internet community
- collaboration between deep tech and industry, innovation and standardisation
- open science principles as well as transversal challenges including gender equality and sustainability.