Intelligent Information Access and Retrieval Systems

Gabriella Pasi
DISCo - Università di Milano Bicocca
pasi@unimib.it

Gloria Bordogna
IREA - CNR
bordogna.g@irea.cnr.it

Abstract

The first edition of the minitrack on "Intelligent Information Access and Retrieval Systems" (IIARS) has been organized within the "Internet and the digital economy" track of HICSS 2023.

We believe that the definition and development of IIARS is nowadays facing new challenges given both to the availability of big data on the Web and to the recent advances and application of machine learning and deep learning methods for Natural Language Processing and Information Retrieval. The minitrack has been conceived as a forum to present and discuss recent proposals and applications in the field of IIARS to trace novel emerging approaches and promising case studies.

Keywords: Information Retrieval models and systems, Question Answering Systems, Knowledge based Systems, Search engines.

1. Introduction

The IIARS minitrack is concerned with the theory, implementation and evaluation of intelligent information access and retrieval models, systems and technologies to bust novel application areas and novel contexts.

It has been conceived with the aim to offer a forum where to discuss both theoretical and practical issues related to intelligent information access and retrieval in emerging contexts and areas, such as big data retrieval, community based retrieval, geographic information retrieval, collaborative search and retrieval, Web information access and retrieval. These areas are strictly related with the artificial intelligence and machine learning.

In fact, the last decade has seen a flourishing of Deep Learning (DL) architectures for natural language texts representation and retrieval, as well as for question answering systems (Agosti et al, 2017).

We are at a turning point in which novel challenges are made possible due to the coupling of big data availability on the Web with DL methods, to boost emerging fields such as context-based, personalized and temporal-based Information Retrieval; crowd-sourced Information analysis and filtering; collaborative search and filtering; community question answering systems, intelligent recommender systems, ecommerce systems with intelligent user Interfaces, intelligent infographics for presentation and exploration of Web search results (Qu et al., 2020; Affolter et al., 2019).

Novel issues have to be considered, such as learning with sparse, heterogeneous, unbalanced, possibly biased and scarce labelled data. Quality assurance and assessment for Information retrieval and filtering are thus of great actuality to secure the veracity of data used for training and testing the models and the fairness of retrieval results (Viviani and Pasi, 2017; Andreasen et al., 2021).

Being the first edition of the track, we collected only three contributions on three different relevant aspects of IIARS: one on metadata use to enable matching between available datasets and data demand in Ecosystems Intelligent Agents; one on the theme of cybersecurity with text analytics based on machine learning, and the third one, which has been accepted for presentation at the conference, entitled: "*On the Uneven Gaps between Different Levels of Graded User Satisfaction in Interactive Information Retrieval Evaluation*"; the papers starts by considering that graded user satisfaction has been frequently employed as a continuous variable in information retrieval evaluation. This has been founded on the unsupported assumption that intervals between adjacent grades are quantitatively equal. The paper tackles the problem of examining the validity of this equal-gap assumption by exploring dynamic perceptual thresholds that trigger grade changes in search evaluation. Specifically, it investigates the extent to which users are sensitive to changes in search efforts and outcomes across different gaps of graded satisfaction by running experiments on four user study datasets. The findings indicate that user satisfaction sensitivity, especially in the case of offline evaluation metrics, changes significantly across gaps in satisfaction scale. Moreover, it outlines that the size and direction of changes in sensitivity vary across study settings, search types, and intentions, especially within "3-5"

URI: https://hdl.handle.net/10125/103140
978-0-9981331-6-4

scale subrange. Therefore, this study is a cornerstone for the user-centered evaluation, and it advances the knowledge of heterogeneity in satisfaction sensitivity to search efforts.

We want to thank the anonymous referees who greatly contributed to assure the quality of the selected paper and the organizers of HICSS for accepting our proposal of the IIARS minitrack.

We do hope it will act as a fertile seed for successive richer editions of the minitrack at HICSS conferences in the coming years.

2. References

Andreasen, T., De Trè, G., Kacprzyk, J., Legind Larsen, H., Bordogna, G., Zadrozny, S. (2021) "Perspectives and views of flexible query answering", *LNCS Vol 12871* Springer Verlag, 1-12.

Affolter, K., Stockinger, K., Bernstein, A. (2019). "A comparative survey of recent natural language interfaces for databases". *The VLDB Journal* 28, 793–819.

Agosti, M., Alonso, O., de Rijke, M., Perego, R. (2017). "Data-driven information retrieval". *SIGIR Forum* 50, 10–14.

Qu, C., Yang, L., Chen, C., Qiu, M., Croft, W., Iyyer, M. (2020). "Open-retrieval conversational question answering, in: *SIGIR 2020 - Proc. of the 43rd Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*, 539–548

Viviani, M., Pasi, G. (2017). "Credibility in social media: opinions, news, and health information — a survey", *WIREs Data Mining Knowl Discov*ery, 7:e1209. doi: 10.1002/widm.1209