

# A Study of Challenges and Obstacles in Biomedical Data Search

Faris M. Abomelha , Ingo Frommholz , Paul Sant

Institute for Research in Applicable Computing, University of Bedfordshire, Luton, UK

## ABSTRACT

The fields of Bioinformatics and Biomedicine are knowledge intensive disciplines where information changes over time due to either the availability of new data or its reanalysis and refinement (e.g., removing errors, adding new annotations), and biological data is being produced at a phenomenal rate. This study has investigated the challenges and obstacles scientists and Bioinformaticians in the Medical Biology field face when working with a variety of databases as well as having to combine heterogeneous structured and unstructured data. A qualitative research methodology (interviews, observation) has been used to examine the main challenges that face information seekers in Biomedicine. Based on the discussion of the results, an information retrieval framework is outlined with the aim of supporting information seekers in Biomedicine in their daily tasks.

## PARTICIPANTS' OBSTACLES

The main obstacles that have arisen from the analysis of the data collected are:

1. There are too many databases to choose from (P1)
2. Duplication of data and missing data (P2)
3. Different formats of data (P3)
4. The need for training to be able to use the tools available (P4)
5. The lack of standards for naming, definition and format (P5)
6. The quality and currency of the data differs from one database to another (P6)
7. Potential of human error with the upload of data to the databases (P7)
8. Lack of documentation for locally developed tools (P8)
9. The constant improvement of current tools and databases is hard to keep up with (P9)

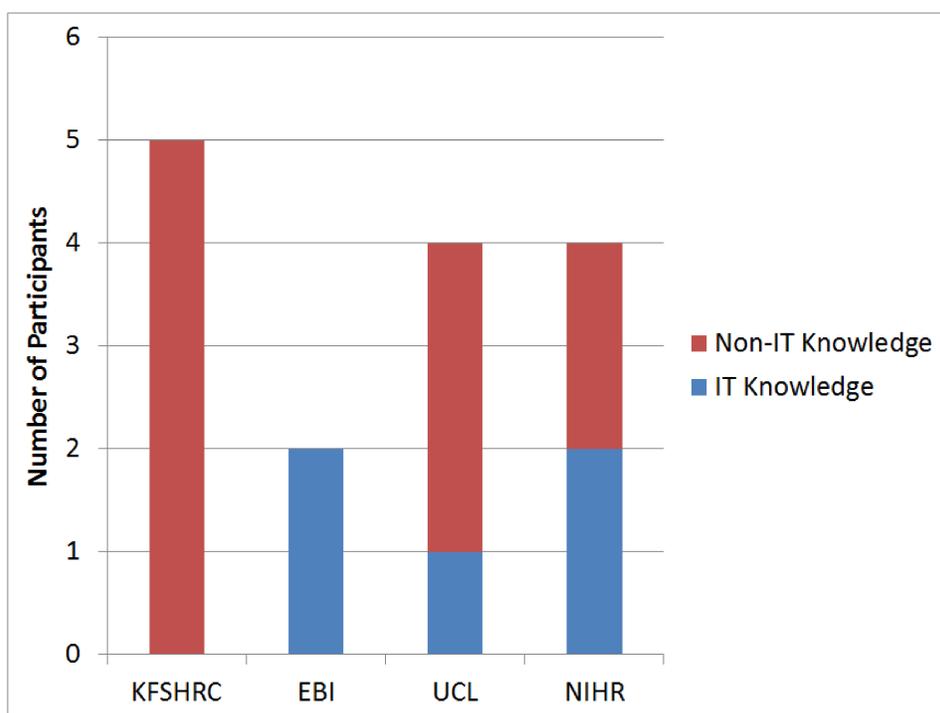


Figure 1: Number of participants

## FRAMEWORK FEATURES

The main features of the proposed IR framework are:

- ❖ Retrieve biomedical data from disparate databases (heterogeneous, textual and non-textual data)
- ❖ Incorporate a focused ranking system for the data (Phred Quality Scoring<sup>1</sup>)
- ❖ Provide the ability for the user to take full control of the records retrieved (edit, update, add)
- ❖ Create a reject list for the user to exclude specific results from appearing again
- ❖ Incorporate a single featured application that will include most of the required databases and tools as tabs
- ❖ Create a straightforward interface that is easy to navigate

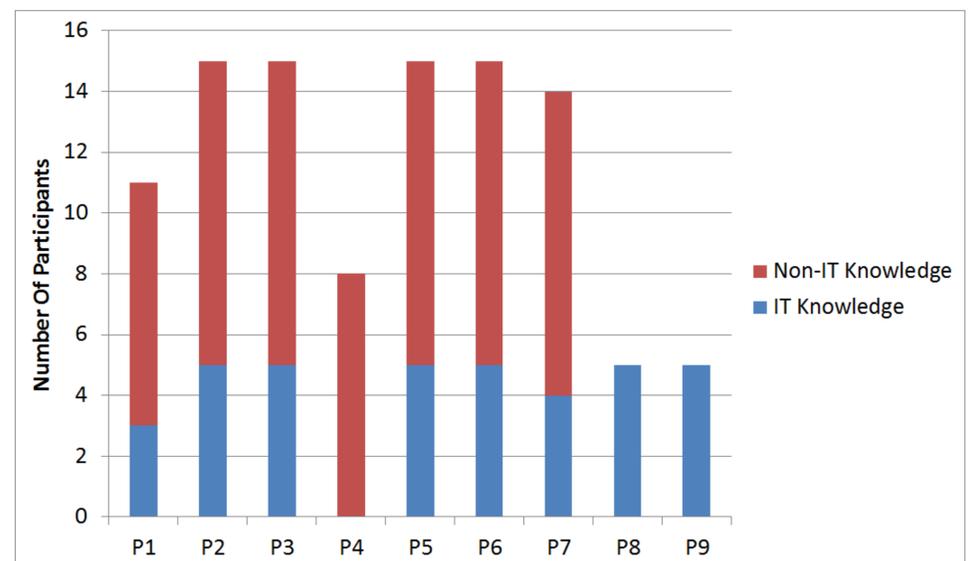


Figure 2: Participants' obstacles

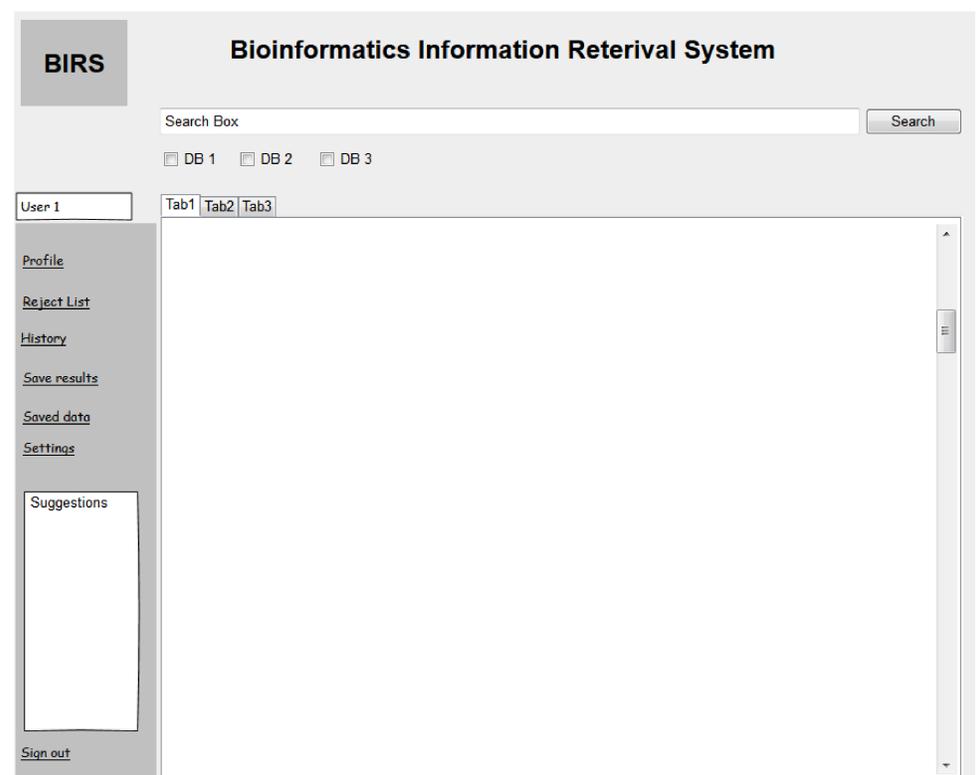


Figure 3: Bioinformatics Information Retrieval System Interface

<sup>1</sup> This is a measure of the quality of the identification of the nucleobases generated by automated DNA sequencing.