

Linking research data with scholarly publications

Teresa K Attwood^{1,2}, Philip McDermott^{1,2}, James Marsh³, Steve R Pettifer², David Thorne³

¹School of Computer Science, University of Manchester, Manchester, UK

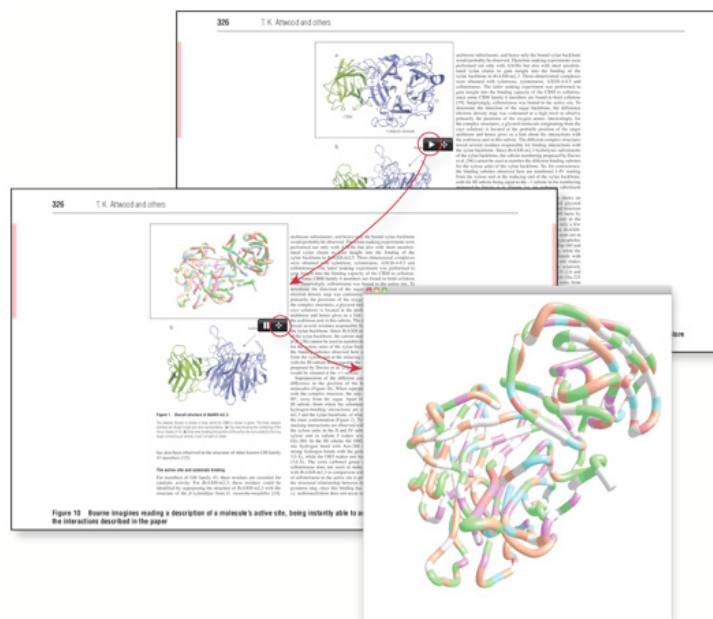
²Faculty of Life Sciences, University of Manchester, Manchester, UK

³School of Chemistry, University of Manchester, Manchester, UK

<http://utopia.cs.man.ac.uk/>

Motivation: in recent decades, a vast gulf has opened between the mass of accumulating research data and the massively expanding literature describing and analysing those data. The problem is not so much data generation per se, but rather, the way in which we've buried the knowledge embodied in those data: there is now so much information available that we simply no longer know what we know, and finding what we want is hard, because the knowledge we seek is often spread across thousands of databases and millions of articles in thousands of journals. The intellectual energy required to search this array of archives, and the time and money this wastes, has prompted the development of new software tools to help link these resources, and ultimately liberate the knowledge that's been systematically trapped within them.

Results: to address some of these issues, we have developed Utopia Documents. Building on Utopia, a suite of semantically integrated protein sequence/structure visualisation and analysis tools (1,2), Utopia Documents is a PDF reader that integrates Utopia's functionality with research articles. The system was piloted in a project with Portland Press to create the Semantic Biochemical Journal (BJ) (3) – in this project, Utopia Documents was used to transform static document features into objects that can be linked, annotated, visualised and analysed interactively, thereby transforming the reading experience and making further analysis from within a PDF file possible for the first time. The Semantic BJ was launched in December 2009 (see www.biochemj.org/bj/424/3/), and Utopia Documents is now being used by BJ editors within their routine publication pipelines. With support from other publishers, and groups like SeqAhead, this new software could also make significant advances towards tighter coupling of NGS literature and data in future.



References

1. Attwood, T.K. et al. (2010) Utopia Documents: linking scholarly literature with research data. *Bioinformatics*, 26, i568-i574.
2. Pettifer, S.R. et al. (2004) UTOPIA - User-friendly Tools for OPERating Informatics Applications. *Comparative and Functional Genomics*, 5, CFG359.
3. Attwood, T.K. et al. (2009) Calling International Rescue - knowledge lost in literature and data landslide! *Biochemical Journal*, 424(3), 317-333.

Relevant Web sites

4. <http://getutopia.com/>
5. <http://utopia.cs.man.ac.uk/utopia/>