

Foreword

It was an afternoon in the early November of 2010 when I sat in front of my laptop thinking on ways to highlight the topic of this book you have in hand. Having been involved in the area of business intelligence for quite some time, I knew that business intelligence over the Web has been a rather neglected topic by the largest part of the research community and not frequently appearing in the mainstream research venues. For me this was a reason good enough for a reference book on the topic, however, I felt I needed a little bit more breadth in my knowledge of the area. As typically happens with people involved in similar tasks, I resorted to the Web (and if you think better about it, this is already an argument by itself) and browsed to Wikipedia's entry for "business intelligence" (in fact, there is an entire category of articles for business intelligence in Wikipedia, but I went right to the homonymous, "main" page). There, I did a simple text search for the word "Web." Surprise: the term was not used in the page at all. A month later, there are a couple of low-key appearances of the term in the same page...

The whole situation is actually mind-provoking, since we do know for sure that the world is rapidly changing towards publishing both data and functionality over the Web. Tim Berners-Lee, in his speech in TED 2009, gave the tone for the future of Internet's usage, by asking the audience to "demand raw data!!" The main idea is that more and more, individuals and government agencies will post data on the Web; analysts can then retrieve the data they deem relevant and work with them. Yet, people in the Information Systems engineering world know that for data to be used by business intelligence tools, we need them to have structure, integrity, compliance to the reference values of the underlying databases, and in general, conformance to the principles regulating the structure and integrity of an Information System (and in particular, a data warehouse). On top of that, even if we can indeed obtain the structure for these Web data, are they trustworthy? Relevant? Available when needed?

At the same time, the Web has already transformed from an "announcement board" where people would post and retrieve information to an environment with two characteristics: (a) a collaboration platform where people share and reuse data in order to work together, and, (b) an automated processing environment where enterprises export functionalities in an automated way in order to conduct business. So, on the one hand, there is a move towards working with collaborative tools and the Semantic Web, and on the other hand, we observe the spread of mashups and software facilities exported as services as the means to speed up and automate business processes.

And where does Business Intelligence fit in such a new and versatile environment? Well... you have to read through the book to find out...

Zorrilla, Mazón, Ferrández, Garrigós, Daniel, and Trujillo have done a great job in compiling a book that fills the gap in the related bibliography with an interesting and comprehensive coverage of the challenges that we face for linking business intelligence and the Web. The book contains a mixture

of research insights, system architectures, discussion of tools, and real-world cases and organizes the discussion of these issues in two areas of coverage. First, the book covers the area of exploitation of data that are already present in the Web for the purposes of business intelligence. The book discusses topics like the management of unstructured text and its combination with super-structured environments like data warehouses and OLAP tools, as well as the management of relevance, freshness, and in general, quality of the Web data with a view to its exploitation via business intelligence tools. The second area involves the automation of data processing and the usage of the Web as a platform for business intelligence including topics like BI-as-a-service, mashups and the Semantic Web for business intelligence, and, collaborative business intelligence.

The book mainly acts as a reference for the state of the art in several areas, including problems and challenges that are not straightforward to be addressed as well as suggestions for paths to follow. The book's primary target is breadth of coverage, with suggestions of the relevant readings for further probing and initial insights for solutions, rather than an in-depth investigation of technical problems in the typical research-oriented fashion. In this sense, the text is easy to follow without losing its interest or significance. In my opinion, the book's primary audience is the interested researcher or practitioner who wants to get a broader view of the environment around the combination of Web and business intelligence, with pointers to the state of the art, as well as the broader challenges that remain open.

Panos Vassiliadis

December 2010, Ioannina