BS-04

existing curve exactly twice. In any event this condition is not fulfilled by the example at the bottom of page 36.

2. The rule that allows a new curve to be introduced (page 42) is not correct; condition 3 should read: 3.(i) if r' is a region of D' which has a counterpart r in D then r' has an X-sequence (is shaded) if and only if r has an X-sequence (is shaded) and (ii) if r' is a region of D' that has no counterpart r in D then r' has no X-sequence; if r' is shaded then r' is a subregion of a region r'' of D' which is shaded and has a shaded counterpart in D.
3. A similar error occurs in the presentation of the rule of unification on page 43. Both errors are reflected in the soundness proofs for the rules.
4. In the hybrid system, no region can correspond to a 'set abstract' of the form $\lambda x \forall y \psi$, yet this is a case considered in Lemma 28 (page 61).
5. The sentence asserted to be satisfiable in the diagram on page 67 is not; one correction could be $\exists x(P(x) \wedge Q(x)) \wedge \neg\exists x \neg(P(x) \vee Q(x))$.

COMPARISONS AND SUMMARY

I am not convinced that either author has produced an ideal account of the *syntax* of their languages. The underlying languages are very large as any differentiable closed curve in the real plane may be used as part of a diagram. The rules for combining curves mean that an arbitrarily large amount of information may be presented in a diagram of a fixed size. This is not consistent with Turing's analysis of the properties that a computable system must possess, so we might be led to conclude that the operations of checking that a diagram is syntactically well formed, or that a proof is valid are not even Turing computable. It is, however, plainly possible to implement systems of diagrammatic inference on a computer, and it would be interesting to develop an approach to syntax which lent itself to such an implementation. Just we are happy to regard the exact shape of a letter as incidental to its syntactic role in a sentence, we could think of the shape of a labelled curve or the position of an x within a region as being incidental to its syntactic role in a diagram. Following this line of thinking would lead to a formulation of the various rules of inference in a more symbolic form, at the risk of obscuring the underlying intuitions behind them.

Neither author attempts to prove the independence of the rules for any of their systems. Hammer's rules for the heterogeneous system are not independent and I strongly suspect that the same holds for his system of Pierce diagrams.

Both accounts require the notion of a counterpart relation, which Shin treats purely abstractly and Hammer approaches via a labelling of the diagrammatic objects. I find Hammer's approach preferable because the counterpart relation is determined by the labelling, which is a suitably concrete syntactic device.

Shin's key contribution is to show that it is possible to prove soundness and completeness results for systems of logic involving visual reasoning, parallelling similar results for languages based on conventional linguistic structures. It might be thought that visual inference rules were simply isomorphic copies of linguistic ones, but this is not the case, because visual perception naturally leads to different kinds of inference. To my mind, this is what makes her results so interesting. To summarise, I quote from her conclusions: 'As long as a system represents information, we should be able to judge whether or not that system is sound and complete. That representation system does not have to be linguistic'. I would recommend her book to anyone interested in diagrammatic reasoning, beginner and expert alike.

Hammer's contribution is to extend one of Shin's results and to apply similar techniques to establish soundness and completeness for a variety of other diagrammatic logics. His notation and proof methods are more consistent with standard approaches in logic, but because of the errors and misprints I would hesitate to recommend it to anyone without a good background in logic.

IST. EL. INF.
BIBLIOTECA
Posiz. ARCHIVIO
BS-04
1998

DR MATT FAIRTLOUGH
University of Sheffield

KAREN SPARCK JONES AND PETER WILLETT (editors) *Readings in Information Retrieval.* Morgan Kaufmann Publishers. (1997) ISBN 1-55860-454-5. $29.95. 589 pp. Softbound.

In a 1983 letter to the members of the ACM SIG on Information Retrieval (IR), its chairman Gerard Salton (to the memory of whom the book here reviewed is aptly dedicated) expressed his concern that, while in the 1960s 'anyone concerned with the automatic processing of data and text seemed to be working in IR', this trend had seemingly reversed in favour of the database community, leaving only 'a small number of active participants in IR work'. Fifteen years later, although databases are still an important topic of research, there is little doubt that the excitement lies elsewhere, and (ironically enough) close to where it was in the 1960s: if a poll were made, asking to name the computer science technology that has had the most profound impact on the general public in the last five years, Internet 'search engines' would probably crush the opposition. Key contributors to the ever-growing success of the Internet, these children and grandchildren of the IR systems that Salton, Sparck Jones and other pioneers have been experimenting with since the 1960s, have arisen an excitement that will no doubt contribute to the success of these *Readings*.

This authoritative work brings together a set of key papers that have contributed in laying the scientific foundations not only of Internet search engines, the conspicuous 'tip of the iceberg' of IR nowadays, but also of the widely installed, if less glitzy, IR systems that are the essential everyday tools of the trade to many organisations. The choice of papers has favoured (a) technical papers that have had a determining influence on the field, rather than review papers;

(b) papers proposing approaches and techniques that have since become established, rather than recent articles that have not yet withstood the test of time; (c) papers that altogether cover IR as a whole and, to a certain extent, 'advanced' information management tasks such as document filtering, categorisation and summarisation.

I especially appreciated the 'categorisation scheme' that was chosen for this book. The papers are subdivided into eight chapters, respectively centred around the history of IR (Chapter 2), the key concepts of IR (Chapter 3), the evaluation of IR systems and techniques (Chapter 4), the formal models that underlie IR systems (Chapter 5), the techniques that are used for implementing them (Chapter 6), important experimental IR systems (Chapter 7), and recent extensions to the basic IR paradigm, including the above-mentioned 'advanced' information management tasks, and the management of new types of documents, such as multimedia (Chapter 8). The final Chapter 9 contains a single, extremely thought-provoking paper discussing the inherent limits of IR. Of special importance, an index of no less than 15 pages is included at the end of the book (not a common feature of previously published article collections).

Each chapter is preceded by an introduction in which the editors present an overall view of the theme of the chapter, put the included articles in the context of this overall view, and provide key bibliographic references integrating those already contained in the articles themselves. These introductory chapters are no routine summaries, but offer to the reader an original synthesis and a stimulating interpretation of the theme of the chapter.

This is an important book, and one that was largely needed. Information retrieval is a discipline that has matured slowly, although steadily, and one in which technological transfer, from research laboratories to the factory floor, has been slower than elsewhere. It is only recently that IR techniques or concepts (e.g. ranked retrieval) that had since long proven valuable in experimental settings, have been incorporated into commercial products. Because of this, many key papers that have determined today's developments, and are thus still essential reading, are scattered through rather old and sometimes hard to find journal volumes and conference proceedings (my estimate is that only a fraction of the key papers included here can be found in libraries with a less than robust information science and library science section). It is extremely appropriate, then, that a well thought-out choice of important papers in this area should have been made available in compact form and at a reasonable price.

Unlike neighbouring disciplines (such as e.g. databases) for which a wide choice of textbooks is available, since the mid-1980s IR has been suffering from a chronic lack of good, comprehensive, and up-to-date textbooks. Although not a textbook, these *Readings* will no doubt be an extremely useful support for lecturers offering IR courses, thanks not only to their balanced choice of material, but also to their excellent organisation into a natural thread that develops from the basic concepts to the recent, cutting-edge applications.

While this book will be definitely appreciated by practitioners of neighbouring disciplines (here I am especially thinking of databases, natural language processing, multimedia, hypertext, library and information science), by graduate students, and by information professionals, to whom it offers a unique and first-hand look at the development of this discipline, it is nothing less than an absolute must for anyone actively involved in IR, either in research or development.

FABRIZIO SEBASTIANI
Istituto di Elaborazione dell'Informazione
Italy