Resolving Redundancy: A Recurring Problem in a Lessons Learned System

John O. Everett
Daniel G. Bobrow
{jeverett, bobrow @parc.xerox.com}

Xerox Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, California, 94304

Abstract

The value of a lessons learned collection depends on how readily the right lesson can be retrieved at the right time. In this paper, we discuss one such collection, from the Eureka system for exchanging tips on photocopier repair. Feedback from Xerox photocopier repair technicians using Eureka indicates that redundancy in the knowledge base is a significant impediment to effective use of the collection. To address this, we are starting a project to develop knowledge-intensive approaches to resolving the redundancies that naturally accumulate in a focused document collection with many authors.

Introduction

The Eureka system is a technician-authored, expert-technician validated set of lessons learned dealing with the repair and maintenance of photocopiers and printers. This paper first describes the origins of this lessons-learned system, and then discusses a significant problem that arises in the use of this successfully fielded system as its knowledge base continues to grow over time, the accumulation of redundant and stale content.

Photocopiers are comprised of computational, optical, electro-mechanical, and chemical subsystems, which renders field diagnosis and repair a challenging task. Xerox has historically spent much money to provide its technical service force with excellent documentation that guides the technician through diagnosis to repair. Unfortunately, the books were so heavy that technicians left them in their cars unless they ran into an unfamiliar or extremely difficult problem. And since the documentation was so expensive to prepare, it was revised very seldom, and hence became out-of-date in crucial places.

To provide a lower-cost and lighter-weight solution with comparable diagnostic capability, a group of researchers at Xerox PARC embarked on a research program to build an automated diagnostic system that would run on a laptop computer. Based on an abstract model of the copier, the system guided technicians through an optimal path to diagnose problems, computing possible probes from probabilities of failures, and expected costs.

A model of one of the more complex copier subsystems was constructed, and a prototype of the system was shown to technicians in the field. They were amazed by what they saw. However, they pointed out that the only problems for which the system would provide guidance were those that arose from known faults; but this is where experienced technicians need the least help. Difficult problems arise from the interaction of multiple factors, such as machine age, local climate, and usage patterns, and not from known single point failures. The subtly nuanced reasoning and knowledge of interactions required for a practical diagnosis system was beyond the state of the art in artificial intelligence.

The PARC team then started to study and work with the technicians to understand their work practices. The technicians themselves pointed out (as documented by Julian Orr in *Talking About Machines*) that war stories exchanged at parts drops and after work often helped them solve their hardest problems. They frequently shared cheat sheets on the current hard problems within their local work groups. PARC researchers suggested that if they had a way of sharing with other more remote communities, they could all do a better job more easily. The PARC team put together some technology integrated with a process designed with the technicians that would help meet their knowledge sharing needs [Bell et al 1997; Bobrow, Cheslow and Whalen, 1999].

This effort evolved into the Eureka system, which today interconnects thousands of widely distributed technicians with a laptop-based repository of technician-authored tips. Technicians periodically log in to update their tips, and to submit new tips to a set of validators. These validators are themselves senior technicians, who attempt to ensure that only high-quality tips are entered into the tip repository. Each tip includes the author's name and contact information, which provides a community incentive for sharing through increased visibility and reputation enhancement. It also encourages responsibility, much like the scientific publishing process. Unlike the scientific publishing process, the validators' names are also provided, since they bear primary responsibility for the quality of the document collection, and continued interaction with them has value to the community.

The Eureka knowledge base is an example of both a lessons learned system, and of a more general class that we call a *focused document collection.* Focused document collections deal with a specific domain (copier repair problems). They are used for specific tasks, and grow relatively slowly. Eureka, for example, contains about 30,000 documents, and is growing by 10-20% a year. It is also a valuable corporate asset, since the use of Eureka is estimated to have saved Xerox roughly $50 million since its inception. Technicians unanimously like the system, and *Knowledge Management World* magazine named Xerox Company of the Year for 1999 based on Eureka.

The Nature of the Problem

As the Eureka document collection continues to grow, the validators who maintain the quality of the knowledge base have found it increasingly difficult to locate potentially redundant tips. This redundancy is problematic because a technician searching for a relevant tip can be overwhelmed by a large number of documents, or confused by similar but not completely consistent ones. Redundancies arise for several reasons. In some cases, several technicians encounter the same problem at about the same time and submit similar tips on how to address it. In other cases, early tips that hypothesize inaccurate diagnoses are corrected by later tips. Validators try to retrieve and take into account earlier tips about the same problem, to minimize redundancy in the database. But relevant tips may not be easily retrieved given the search tools available, and the validators are under time pressure, and may not always be able to check thoroughly

Finally, different authors emphasize different aspects of the problem. For example, some of the tips provide explanations of the issue, others focus on workarounds (what to do until field engineering comes up with a new part to replace one that is unexpectedly failing), and yet others provide only a terse description of the symptoms, how to confirm the problem, and what field engineering specifies as the official fix. We have found many different relationships among similar tips, including contradictions, elaborations, alternative hypotheses, and hints on how to work within the Xerox corporate environment.

Figure 1 shows three similar (expurgated) tips for a mid-volume photocopier. The first of the three tips postulates a problem in a particular subsystem that is expensive to manufacture and time-consuming to replace. Naturally, field engineering was keen to pinpoint the root cause of problems with this subsystem, so the tip encourages technicians to return defective ones for examination. In the middle tip, however, there is a note toward the end, to the effect that this subsystem is not the problem. Apparently, moving the wiring harness in order to remove the subsystem caused an intermittent connection to work again for a time, leading technicians to believe the subsystem was at fault, and in fact, 95% of those returned were fine. The final tip tersely summarizes the situation, as "bimetallic corrosion" (i.e., tin pins going into gold plugs), without any reference to the prior tips.

Clearly, having all three tips in the knowledge base is problematic, especially if a technician were only to look at the first one. Although the first tip should probably be deleted, the second tip contains useful elaborations on the situation that may be of use to new technicians.

There is substantial evidence that this problem is pervasive. There are an increasing number of active users of the data base, they uniformly like the system, but requests for tools to manage the redundancy have increased as well. These requests have also been pressed upon us from many different validators, and over the past eighteen months have taken on a tone of increased urgency. Our own analyses support the field's contention of significant redundancies in the corpus. A manual analysis of 650 tips turned up 15 pairs (or triples in a couple of cases) of conceptually similar tips. Because technicians tend to use coarse search terms (e.g., "fuser"), search results tend to be large and contain any similar tips that exist in the database

To maintain the value of the Eureka system as it continues to scale up, it would be useful to automate (at least partially) the identification and resolution of such redundancies. Whereas people excel at making fine distinctions among small numbers of natural language documents, computers excel at processing larger numbers of document There is a clear need to develop tools that play to both of these strengths, by automating the identification of similar tips, but leaving the resolution of redundancy up to the maintainers of the system.

There are two obvious approaches to handle this problem, neither of which works in practice. The first is to ask the technicians to characterize the tip during entry, by selecting keyword terms etc, or even having editors "clean up" their work. This would make the tips indistinguishable from other corporate documentation. Of all the information sources available to the technicians, only Eureka is completely under their control. A general refusal to select terms from pull-down lists of metadata intended to more accurately characterize and index the tips arises from the same sentiment.

The second approach is to use standard information retrieval methods, such as word overlap, perhaps extended by latent semantic analysis. We have tried this, and it is of limited value for this corpus, which has characteristics that render it generally impervious to statistical analyses.

The average tip is terse (a little over a hundred words) and liberally sprinkled with jargon, misspellings, and creative departures from generally accepted punctuation and grammar. Similar tips may overlap by as few as fourteen words. Although tips are structured in Problem, Cause, and Solution sections, authors vary in their use of this convention.

Our Approach

The Eureka system is similar in some respects to a case-based reasoning system, but there are significant differences. Although tips may be construed as cases, they are authored by technicians, not knowledge engineers, and they adhere to no common format or metadata convention. The associated search engine, SearchLite, is a full-text

Tip 36321

Problem: Subsystem failing intermittently with error code E-12

Cause: Unknown

Solution: If the subsystem responds to the following procedure, replace it and send the defective one to field engineering.

Tip 41023

Problem: Error code E-12

Cause: . . .The tin plating on the wire harness pins oxidizes and creates a poor connection.

Solution: ..."slide" the connectors on and off. This will scrape the pins and provide a better contact. Long Term: Harnesses with gold plated pins will be available... The subsystem is usually NOT the problem

Tip 46248

Problem: Intermittent faults in subsystem

Cause: Bi-metallic corrosion

Solution: A new kit containing harness with gold plated pins to replace old harness is now in parts supply.

Figure 1: Evolution of Knowledge in Eureka

keyword-based system tuned to the particularities of the domain. For example, it recognizes fault codes and part numbers as such, which improves the accuracy of its tokenizer, but it does not rely on explicitly encoded domain knowledge.

Several attempts to have technicians structure their entries (e.g., by tagging them with metadata from pull-down menus) ended in failure. To the technicians, pressed for time and often writing tips after hours, the advantages of natural language text in terms of rapid entry and ease of use were so great that they systematically refused to use other interfaces.

We believe that unstructured but focused document collections such as Eureka present a large class of interesting and relevant problems that will yield to knowledge-intensive methods. Finding the resources for skilled knowledge engineers to maintain such collections is often difficult in organizations under budgetary pressures, so automating aspects of this maintenance may prove quite valuable.

In contrast to much of the on-going research in case-based reasoning, the field most closely allied to this work, our research is deliberately challenging the conventional wisdom that it is not possible to extract useful representations of the underlying contents from unstructured text via natural language processing (NLP). Limited task-focused dialog understanding systems have existed since the 70's [Bobrow et al 1977], and are now incorporated into commercial speech understanding systems for limited interfaces. However, difficulties encountered in providing a fuller understanding of a significant domain seemed insurmountable at the time. We believe, however, that enough progress has been made in the underlying theory and technologies to warrant another principled attack on this understanding problem.

We intend to apply deep parsing and knowledge representation techniques to the problem of maintaining the quality of the Eureka knowledge base for three reasons. First, we believe that recent technological advances have brought us to the verge of making such techniques commercially feasible, especially when applied to document collections focused on a particular domain, such as photocopier repair. Recent implementations of Lexical Functional Grammar theory [Kaplan and Bresnan 1982] have demonstrated cubic performance in parsing a wide and increasing range of sentences, with small enough constants that an average fifty word sentence can be parsed in a second or less.

Second, the brevity of the tips and the high degree of common knowledge assumed reduces the effectiveness of statistical approaches, while making it more likely that the contents can be represented in terms from a limited domain. We expect that techniques for indexing, retrieval, and representation developed by researchers in analogical reasoning, statistical language processing, and case-based reasoning to be complementary. Our hope is that we can achieve higher accuracy with deep representations than would be possible with current state of the art techniques.

Third, we believe that a knowledge-intensive approach will provide the basis for automating other document-collection tasks. We will be investigating the viability of creating composite documents from parts of existing tips, in a process we call *knowledge fusion*.

Related Work

There is an extensive literature on case-based reasoning systems, some of which touches directly on the issues of redundancy and knowledge base management that we are addressing. Time and space preclude a thorough review of this literature here, so we can only highlight relevant research here. The primary difference between CBR work and our research is that CBR systems generally presume the existence of an infrastructure for encoding cases, whereas we are starting with a large unstructured corpus.

Racine and Yang [1997] address the issue of maintaining a largely unstructured corpus of cases via basic information retrieval methods, such as fuzzy string matching. Although we have found some instances of what could only be multiple edited versions of the same text, for the most part we are looking for conceptual similarities that tend to manifest in texts that have surprisingly little overlap in words used.

Mario Lenz has done some relevant work in textual CBR, in particular defining a layered architecture that includes keyword, phrase (e.g., noun phrase), thesaurus, glossary, feature value, domain structure, and information extraction layers [Lenz 1998]. He and others have implemented several systems based on this architecture, including FALLQ [Lenz and Burkhard 1997] and Experience-Book [Kunze and Hübner, 1998], and a system for Siemens called SIMATIC Knowledge Manager [www.ad.siemens.de]. An ablation study [Lenz 1998] shows little degradation in performance if the information extraction layer is not used, but significant degradation if domain structure is not utilized, and a major loss in recall and precision if domain-specific phrases aren't used.

FAQFinder [Burke et al., 1997] attempts to match a query to a particular FAQ (set of frequently asked questions about a topic) and extract from that FAQ a useful response. In contrast to other research here, this system does no a priori structuring of the FAQ corpora, and in this regard is similar to our research. FAQFinder uses a two-stage retrieval mechanism, in which the first stage is a statistical similarity metric on the terms of the query and a shallow representation of lexical semantics utilizing WordNet [Miller, 1995]. This enables the system to match "ex-spouse" with "ex-husband," but it fails to capture causal relations, such as the relation between getting a divorce and having an ex-spouse. The authors point out that developing such a representation for the entirety of the USENET corpora would not be feasible. This supports our strategy of limiting our initial efforts to focused document collections, where we have a reasonable chance of developing sufficient representations for our task. An ablation study shows that the semantic and statistical methods both contribute independently to recall, as their combination produces a higher recall (67%) than either used alone (55% and 58%, respectively).

SPIRE combines information retrieval (IR) methods with case-based reasoning to help users formulate more effective queries [Daniels and Rissland, 1997]. SPIRE starts by retrieving a small set of cases similar to the user's inquiry, then feeds this to an IR search engine, which extracts terms from these cases to form a query, which is run against the target corpus to produce a ranked retrieval set. SPIRE then locates relevant passages within this set by generating queries from actual text excerpts (marked as relevant by a domain expert during preparation of the knowledge base). The user may add relevant excerpts to the case-base from the retrieval set, thus continuing the knowledge acquisition process during actual use.

Information extraction [Riloff and Lehnert, 1994] is another area of relevant research that may be considered as a mid-point between word-based approaches and deep natural language parsing. By limiting the goal of the system to extracting specific types of information, the natural language processing problem is considerably simplified. Desired information is encoded in a frame that specifies roles, enabling conditions, and constraints.

Leake and Wilson [1998] have proposed a framework for ongoing case-base maintenance, which they define as "policies for revising the organization or contents...of the case-base in order to facilitate future reasoning for a particular set of performance objectives." The dimensions of this framework include data collection on case usage (none, snap-shot, or trend), timing of maintenance activities (periodic, conditional, or ad hoc), and scope of changes to be made (broad or narrow).

This line of research is important to our project for two reasons; first there is an extensive body of literature on CBR maintenance techniques which should be directly applicable to our research when we have reached the point of automatically generating reasonable conceptual representations. For example, Smyth and Keane [1995] have investigated methods for determining which cases to delete without impacting performance, many researchers [e.g., Domingos 1995, Aha, Kibler & Albert 1991] have presented methods for automatically generalizing cases, and [Aha and Breslow, 1997] present a method for revising case indices.

Second, we need to develop a maintenance policy that respects the current work practices and resource constraints of the Eureka validators. Leake and Wilson stress the importance of setting a maintenance policy with respect to a set of performance objectives, which is a key consideration for our project. Watson [1997] provides guidelines for human maintenance of CBR systems.

Project Status

This project is just getting started. We have to date hand-coded knowledge representations for one pair of similar tips, to serve as a benchmark for system development. We hope by year-end to show this working all the way through for a small number of tips from this domain. While it remains to be seen whether or not our approach will produce systems with demonstrably better performance, we believe that a careful reexamination of the conventional wisdom in this area is in order, and has the potential to produce an exciting new direction in symbolic AI research.

References

Aha, D.; Kibler, D.; and Albert, M. 1991. Instance-based learning algorithms. *Machine Learning* 6:37-66.

Aha, D., and Breslow, L. 1997. Refining conversational case libraries. In *Proceedings of the Second International Conference on Case-Based Reasoning*, 267-278. Berlin: Springer Verlag.

Bell, David, Daniel G. Bobrow, Olivier Raiman, Mark Shirley. Dynamic Documents and Distributed Processes: Building on local knowledge in field service. In *Information and Process Integration in Enterprises: Rethinking Documents*, eds. Wayakama, T et al. Kluwer Academic Publishers, Norwell MA 1997

Bobrow, Daniel G., Robert Cheslow, Jack Whalen. Community Knowledge Sharing in Practice. Submitted to *Organization Science* 1999

Bobrow, Daniel G., Ronald M. Kaplan, Martin Kay, Donald A. Norman, Henry Thompson, Henry Winograd GUS, a frame-driven dialog system *Artificial Intelligence*, 8 (2) (1977) pp. 155-173

Burke, R., K. Hammond, V. Kulyukin, S. Lytinen, N. Tomuro, and S. Schoenberg. Question Answering from Frequently Asked Question Files. *AI Magazine*, 18(2):57-66, 1997.

Daniels, J.J.; Rissland, E.L. Integrating IR and CBR to locate relevant texts and passages. *Proceedings of the 8th International Workshop on Database and Expert Systems Applications* (DEXA) 1997.

Domingos, P. 1995. Rule induction and instance-based learning. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, pp. 1226-1232. San Francisco, CA: Morgan Kaufmann.

Kaplan, Ronald M. and Joan Bresnan. 1982. Lexical-Functional Grammar: A formal system for grammatical representation. In *The Mental Representation of Grammatical Relations*, ed. Joan Bresnan. The MIT Press, Cambridge, MA, pages 173-281. Reprinted in *Formal Issues in Lexical-Functional Grammar,* eds. Mary Dalrymple, Ronald M. Kaplan, John Maxwell, and Annie Zaenen, 29-130. Stanford: Center for the Study of Language and Information. 1995.

Kunze, M. and A. Hübner. CBR on Semi-structured Documents: The Experience-Book and the FAllQ Project. In *Proceedings 6th German Workshop on CBR*, 1998.

Leake, D.B. and D. C. Wilson. Advances in Case-Based Reasoning: *Proceedings of EWCBR-98,* Springer-Verlag, Berlin

Lenz, M. Defining Knowledge Layers for Textual Case-Based Reasoning. In *Advances in Case-Based Reasoning, Lecture Notes in Artificial Intelligence 1488*, eds. B. Smyth and P. Cunningham, pp 298-309. Springer Verlag, 1998.

Lenz, M. and H.-D. Burkhard. CBR for Document Retrieval - The FAllQ Project. . In *Case-Based Reasoning Research and Development, Proceedings ICCBR-97, Lecture Notes in Artificial Intelligence, 1266.* eds., D. Leake and E. Plaza, pages 84-93. Springer Verlag, 1997

Miller, G. A. 1995. WordNet: A lexical database for English. *Communications of the ACM* 38(11):39-41.

Racine, K., and Yang, Q. 1997. Maintaining unstructured case bases. In *Proceedings of the Second International Conference on Case-Based Reasoning*, 553-564. Berlin: Springer Verlag.

Riloff, E., and Lehnert, W. 1994. Information extraction as a basis for high-precision text classification. *ACM Transactions on Information Systems* 12(3):296-333.

Watson, I. D. *Applying Case-Based Reasoning : Techniques for Enterprise Systems.* Morgan Kaufmann, 1997.