Generating Reports from Case-Based Knowledge Artifacts

Rosina O. Weber, Sidath Gunawardena, Jason M. Proctor

College of Information Science & Technology, The iSchool at Drexel University, Philadelphia, PA 19104

{rw37, sg349, jp338}@drexel.edu

Abstract

Knowledge artifacts are units retained in repository-based knowledge management systems (KMS). When adopting cases as the representation for knowledge artifacts in KMS, one of the benefits is the automated construction of reports.

Introduction

Repository-based knowledge management (KM) systems retain a memory of artifacts that are to be shared among members of the community whose knowledge the system targets to manage. The basic concept of reusing previous experiences in case-based reasoning (CBR) is quite similar, suggesting it as a suitable methodology for implementing KM systems. Althoff and Weber (2005) further discuss and illustrate KM and CBR affinities.

The context of this paper is a repository-based KM system developed for a community of scientists (Weber et al. 2006). The knowledge artifacts submitted by members of this community are scientific contributions. Their knowledge artifacts can also be used to populate annual reports. Therefore, we considered the perspective of automatically drafting reports as a valuable compensation for the time they spend contributing artifacts.

Building reports such as those typically written by research scientists is a natural language generation (NLG) task (Reiter and Dale 1997). Our task is more specifically of drafting as contributors will revise the reports.

Aha et al. (2001; Weber and Aha 2003) proposed a representation for knowledge artifacts based on cases that was shown to improve plan quality. The case representation they proposed inspired the work discussed in this paper, which presents another benefit of using such a case representation for KM systems. The benefit is this paper's major contribution, to demonstrate how to generate automated reports from knowledge artifacts represented as cases. These reports can help research scientists by performing the ground work of reusing what they have already submitted to the repository.

Problem Context

The knowledge repository (KR) we discuss in this paper was designed for a community of scientists. They are investigators for the Center for Advancing Microbial Risk Assessment (CAMRA). The system is the CAMRA KR.

The resulting knowledge artifacts based on cases are called learning units (LUs). The categories Complete, Progress, and Read are described next. Examples of these categories covering the complete process involved in sharing research activities are given in Figure 1.

Complete is designated for completed research activities. The contributions described under this category are ready to be published in some medium (e.g., a conference paper).

Progress refers to the preliminary stages of a research activity. It is distinguished from the first category in its reuse elements. The indexing elements are basically the same; reuse elements are different. In this stage of a research activity, contributions are not yet available. Therefore, we ask for a hypothesis and the steps to validate it, rather than the knowledge learned.

Read is the label of LUs that describe things learned from the literature. What is learned is supported by references.

From Learning Units to Report Generation

Different categories for LUs trace the steps performed in a research activity through associations. For example, Complete LUs are preceded by Progress units; ongoing research may be reflected in a series of Progress LUs.

LUs include two indexing components that describe and specify a *research activity*. The reuse elements represent the *knowledge* learned and its *validation*. The principle of the report is to start by describing and

Read	Progress	Complete
Research activity. Identifying a microbe	**Research activity.** Identifying microbe fate	**Research activity.** Identifying microbe fate
Specifics: Environment is water	**Specifics:** Agent is Microbe X	**Specifics:** Agent is Microbe X
Validation: Smith J 2006. Introduction to Microbe X. Journal of Bugs 21(8):56-72	**Steps to Validate:** We will freeze samples to find survival rate at various temperatures	**Results of Experiments:** 95% of sample population died after 24 hs in 0°C
Knowledge: A new microbe, X, was identified. Fate and transport are unknown	**Hypothesis**: Microbe X dies after prolonged exposure to freezing temperatures	**Knowledge:** Microbe X dies after being exposed to 0°C for 24 hs

Figure 1. Examples of learning units (LUs)

specifying the research activity, i.e., the indexing. Next, the report provides the steps to *validate* the *knowledge*. Finally, the *results of experiments* are presented in support of the learned knowledge and the text is completed with the *knowledge* learned, i.e., reuse elements.

Table 1. Templates for specific associations between LUs

Template	Association between learning units
Template 1	One Progress Unit, second -Progress Unit
Template 2	One Progress Unit, one Complete Unit
Template 3	One Progress Unit, second Progress Unit
Template 4	One Progress Unit, one Complete Unit, second Progress Unit
Template 5	One Progress Unit, one Complete Unit, multiple Progress Units

The first step is to identify the LUs to include in the report. The second step is to find LUs that are linked by associations, which are used to select a *paragraph template*. This is analogous to document planning (Reiter, Sripada, and Robertson 2003). Paragraph templates have the scope of units that have categorical associations. A *paragraph template* determines the structure and number of paragraphs of the report. The selection of the right template is done by the associations as shown in Table 1.

The third step is analogous to micro-planning (Reiter, Sripada, and Robertson 2003). *Sentence templates* are selected to determine the final text by suggesting the wording surrounding the contents of a LU's field. The grammatical verification component of the surface realization (Reiter, Sripada, and Robertson 2003) process is to be done by humans when they revise the reports.

Study

In order to examine the further design of our reporting capability, we submitted our application to all first year's LUs in the CAMRA KR of the types Progress and Complete. The output of the application comprises the main section of the report.

To examine the suitability of the automated reports, we compared the number of topics in the investigators' reports to the number contained in the automated reports (Table 2). We distinguish topics by a clear shift into a different topic, as interpreted by a non-expert.

Table 2. Comparison between manual and automated reports

	Project	1	2	3	4	5	Total	Avg.
No.Topics	Investigator Reports	3	6	5	2	6	22	4.4
	Automated Reports	2	3	3	1	5	14	2.8

The reasons why automated reports include 36% less topics than investigators' reports still need to be identified. It is relevant to note that the reports submitted by investigators were to be summarized by the center's directors before submission to the funding agency. The value of the LUs was recognized when only the contents of LUs were used in the final summary.

Conclusions and Future Work

We showed how to automatically generate reports from case-based knowledge artifacts. The resulting reports in our preliminary study show the method is promising. These reports will be offered to the members of the CAMRA community as a compensation for the time they spend contributing LUs. We expect that this byproduct becomes a motivator of contributions to the CAMRA KM.

The contribution of this paper is to demonstrate how indexing and reuse elements of case-based knowledge artifacts can be used to build automated reports in different communities. The validation of this capability will be done with a qualitative study.

Moreover, the same capabilities could be used to report ongoing efforts of the group. The variety of templates using Read and Progress units can be used to inform project and center leaders of the group's progress.

Acknowledgements

This work is supported in part by the U.S. EPA-Science to Achieve Results Program and the U.S. Department of Homeland Security Programs, Grant # R83236201.

References

Aha, D. W., Weber, R. O., Muñoz-Avila, H., Breslow, L. A. & Gupta, K. 2001. Bridging the Lesson Distribution Gap. In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence*, 987-992. San Mateo, CA: Morgan Kaufmann.

Althoff, K.-D., and Weber, R. O. 2005. Knowledge Management. Case-Based Reasoning. *Knowledge Engineering Review* 20(3): 305–310.

Reiter, E. and Dale, R. 1997. Building Applied Natural-Language Generation Systems. *Journal of Natural-Language Engineering* 3:57-87

Reiter, E., Sripada, S., and Robertson, R. 2003. Acquiring Correct Knowledge for Natural Language Generation. *Journal of Artificial Intelligence Research* 18:491-516.

Weber, R. and Aha, D.W. 2003. Intelligent Delivery of Military Lessons Learned. *Decision Support Systems* 34(3):287-304.

Weber, R. O., Morelli, M. L., Atwood, M. E. and Proctor, J. M. 2006. Designing a Knowledge Management Approach for the CAMRA Community of Science. In *Proceeding of the Sixth International Conference Practical Aspects of Knowledge Management*, *Lecture Notes in Artificial Intelligence* 4333, 315–325. Berlin Heidelberg: Springer-Verlag.