

Two-phased Knowledge Formalisation for Hydrometallurgical Gold Ore process recommendation and validation

Christian Severin Sauer · Lotta Rintala · Thomas Roth-Berghofer

Received: date / Accepted: date

Abstract This paper describes an approach to externalising and formalising expert knowledge involved in the design and evaluation of hydrometallurgical process chains for gold ore treatment. The objective of this knowledge formalisation effort is to create a case-based reasoning application for recommending and validating a treatment process of gold ores. We describe a twofold approach to formalise the necessary knowledge. First, formalising human expert knowledge about gold mining situations enables the retrieval of similar mining contexts and respective process chains, based on prospecting data gathered from a potential gold mining site. The second aspect of our approach formalises empirical knowledge on hydrometallurgical treatments. The latter will enable us to evaluate and, where needed, redesign the process chain that was recommended by the first aspect of our approach. The main problems

with the formalisation of knowledge in the gold ore refinement domain are the diversity and the amount of parameters used in literature and by experts to describe a mining context. We demonstrate how similarity knowledge was used to formalise literature knowledge. The evaluation of data gathered from experiments with an initial prototype workflow recommender, *Auric Adviser*, provides promising results.

Keywords Knowledge formalisation · Hydrometallurgy · Case-based reasoning

1 Introduction

Nowadays rich gold ores that can easily be processed with simple metallurgical processes like direct smelting are getting rare. This situation leads to the (re)evaluation of many less rich and difficult to process gold ore deposits, considered too cost intensive for mining before. Refractory, or in other words difficult to process, gold ores are gold ores that in general require complex and cost intensive processes to extract a comparatively small amount of gold from a large volume of ore. Thus the large scale processes involved in the processing of such ores are to be planned carefully to avoid failed investments in ore processing facilities either not adequate for the ore mined at the mining site or not efficient enough to extract sufficient quantities of gold and thus sufficient revenues.

A key problem in today's prospecting for gold mines and in their planning is the estimation of the costs in-

Christian Severin Sauer
School of Computing and Technology
University of West London, United Kingdom
Tel.: +44-(0)20-8280-0250
E-mail: christian.sauer@uwl.ac.uk

Lotta Rintala
School of Chemical Technology
Research Group of Corrosion and Hydrometallurgy
Tel.: +358-50-460-3280
E-mail: lotta.rintala@aalto.fi

Thomas Roth-Berghofer
School of Computing and Technology
University of West London, United Kingdom
Tel.: +44-(0)20-8231-2719
E-mail: thomas.roth-berghofer@uwl.ac.uk

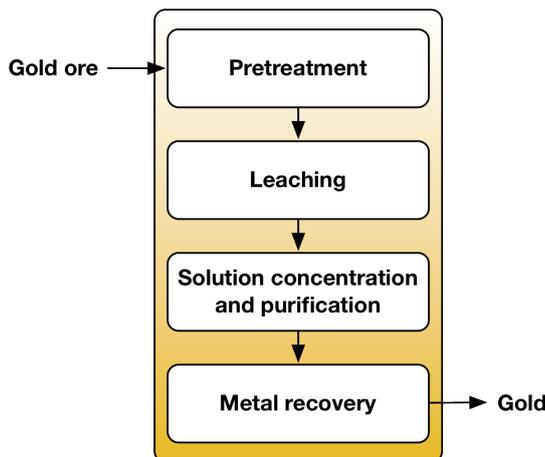


Fig. 1 Basic hydrometallurgical process chain

involved, not only in the mining of the ore but mainly in its processing costs given, for example, by the use of certain chemicals. Additional important factors are the ore throughput capacity to reach a necessary volume of ore to be processed per day as well as the amount of gold to be extracted from this ore. An early and exact estimate of these costs and factors allows excluding early on potential mining operations that are not cost efficient and helps to speed up planning of worthwhile mining operations by reusing process knowledge previously employed in successful mining operations.

In this paper we examine the elicitation, formalisation and re-use of expert knowledge about gold mining situations that operate on refractory gold ores, hydrometallurgical process design for processing these ores and empirical knowledge focused on the diverse hydrometallurgical processes involved in the process chains.

Hydrometallurgy is a field of science which studies the recovery of metals from ores by using aqueous chemistry. A typical hydrometallurgical process chain is illustrated in Figure 1. When analysing or designing hydrometallurgical processes, a process chain is typically considered to be composed out of smaller stages or single process steps, we call treatments. For example, a commonly used leaching technique for gold extraction is cyanide leaching. However when it is a question of refractory gold ores, for example, simple cyanidation might not be effectively used. In such ores, gold is encapsulated inside a host mineral and thus cannot be reached by the leaching agent. Thus the host mineral

must be broken using pre-treatment processes to liberate the gold, before cyanide leaching.

Depending on the level of the refractoriness of the gold ore and the mineralogical characteristics of the ore, a variety of alternative processing routes exist. For example an expert designing a process chain could face the need to choose between oxidising processes, such as roasting, pressure oxidation or bacterial leaching. Typically the design of process chains relies on laboratory tests, which are time consuming and thus cost intensive. Today there already is a large amount of published work available, detailing on experiments about gold extraction from ore. Additionally on-going research is constantly adding to this knowledge. If a process chain designer could easily re-use this existing knowledge more efficiently than manually researching on it, the needed experiments would be selected more precisely or even be made redundant. Furthermore bench and pilot scale experiments, which are testing experiments on a designed process chain before scaling them to production size in a mining operation, would also be achieved more rapidly. Thus the effort for the process chain design could be significantly reduced. The work presented in this paper demonstrates an approach at formalising the existing knowledge on hydrometallurgical process chain design, using CBR and thus make it readily available for re-use by the process chain designer.

The knowledge involved in estimating a mining situation from prospecting data and designing a process chain for the specific gold ore expected in this potential mining context is highly encoded. Usually a few experts in the domain are consulted to give their experience-based estimate of a prospected ore deposit and it being worthwhile mining or not. Additionally there exists a great amount of empirical knowledge, mainly from hydrometallurgical experiments on single process steps or treatments. This knowledge about specific treatments is mainly encoded in scientific publications on specific hydrometallurgical treatments.

In this paper we demonstrate our twofold approach. The first aspect or step in our approach is to reuse knowledge about gold mining situations to retrieve similar mining contexts based upon prospecting data gathered from a potential gold mining site. The second step of our approach is then to evaluate the recommended process chain automatically and recommend a re-design of the process chain where necessary.

In this paper we describe how the knowledge from existing mining contexts was formalised and is used in the *Auric Adviser* workflow recommender software. We further show how *Auric Adviser* recommends on the best process chain to be used in hydrometallurgical treatment of rich and refractory gold ores in a potential mining project. We formalised the knowledge we elicited from human experts and hydrometallurgical publications to re-use it for our case-based reasoning (CBR) approach. Case-based reasoning is an established artificial intelligence technique. It allows for the versatile handling of incomplete and fuzzy knowledge.

The rest of the paper is structured as follows. We interlink our approach with related work on hydrometallurgy, workflow recommendation and knowledge representation in the following section. In Section 3 we survey the knowledge sources targeted in Section 3.2 and introduce our use of the different knowledge containers of CBR to provide the captured knowledge in Section 3.3 and review the resulting knowledge model in Section 3.4. Following a brief introduction to the *Auric Adviser* software in Section 4 we detail on our experiments in Section 5. We then introduce our ongoing work on our second knowledge model for individual treatment recommendation in Section 6, detailing on the sources and structuring of the knowledge in Sections 6.1 and 6.2. The paper then evaluates and discusses the performance of our knowledge model and the *Auric Adviser* (Section 7). A summary and outlook on future aspects of our work concludes the paper.

2 Background and Related Work

Recommending workflows is a research area in CBR (cf. [7,10]). CBR has been used successfully in a number of workflow recommender (cf. [9]). The approach to, at least semi-automatise, the design of mining facilities and their accompanying ore process facilities and processing chains is also an established area of research. As, for example, the work of Torres et al. [16,17] on their IntelliGold system, has shown an approach to use decision rules and fuzzy sets to recommend ore processing workflows and calculate an estimate of the associated costs to establish such a process chain. However after an initial small set of case studies this approach seems to have not been followed further. In our view the complexity and variety of the knowledge in the domain of gold mining and refractory gold ore refining is a factor

that ultimately could not be covered by a rule based system alone.

Picking up on the initial work of Torres et al. we therefore present our new approach to ore processing workflow recommendation. We were confident in the decision to use CBR as we have already successfully used CBR and its versatile use of similarity based retrieval, in the formalisation of knowledge in the highly complex and encoded domain of audio mixing and subsequently implemented an efficient audio mixing workflow recommender [15]. Additionally the use of the CBR methodology is documented as successful in the field of chemical engineering. It is used, for example, for the separation process synthesis and as a method for combinatorial mixer equipment design from parts and the development of feasible separation process sequences and separation process structures for wood processing [11].

3 Knowledge Formalisation Approach for use in CBR

As outlined we used case-based reasoning in our approach to reuse the elicited and formalised knowledge of mining experts and empirical knowledge on hydrometallurgical treatments. We chose CBR as a suitable methodology [18] for our task as it is able to handle the inherent vagueness and broad variety of the knowledge present in our domain of interest.

CBR's suitability for the knowledge formalisation in our domain is backed up by a variety of CBR applications in similarly demanding domains ranging from chemical process design to music composition [2,12]. The manner in which we employed CBR for our knowledge formalisation is the subjects of this section. For the purpose of modelling and testing the knowledge of our system we used *myCBRWorkbench*¹ in its latest version 3.0. We then developed the java-based application, *Auric Adviser*, with the *myCBRSDK*.

Broadly speaking CBR deals with storing/retrieving as well as re-using/adapting experience. It mimics the human approach of re-using problem-solving experience encoded in cases. If we encounter a new problem we most often try to remember a similar problem we solved in the past. We try to match the problem description of the problem at hand with problems we have encountered and solved in the past. The problem description

¹ <http://mycbr-project.net>

and its solution description can be seen as an episode of experience. Cases, in CBR, are problem/solution pairs. Facing a current problem we recall matching past problems and adapt their solution to our current situation [1].

To further formalise this similarity towards human memory, a CBR-system consists mainly of a case base which is determined as a collection of cases. As mentioned a case is then formally described as consisting of at least a problem description and a solution description [1, 6]. A formal representation of a case base is thus given by the following:

[[problem1, solution1], [problem2, solution2], ..., [problem n, solution n]] : case base

The classic process model of CBR, CBR-cycle [1], starts with a new problem or query being considered as a new case. This initial case is consisting of the problem description of the current problem presented to the CBR system. Then, during the Retrieve step, the case-base is searched for the most similar case to the new case. To determine the similarity between the query case and a given case in the case base, similarity measures are used. These similarity measures determine the relevance of attributes for a case and the similarity between different values of one attribute. This ability to ‘compare’ the current problem (case) with all stored problems (cases) allows for the retrieval of either an exact matching problem or at least the retrieval of the most similar problem (case) known to the CBR system so far.

Having retrieved a case the CBR cycle enters its Reuse step. During the Reuse step the solution part of the retrieved case is tried to be applied to the current problem at hand. If the system retrieved an exact matching test the reuse is the simple application of the unchanged solution part of the retrieved, exact matching, case. If however the retrieved case is only the most similar case to the query case the solution has to be adapted. This adaptation can be achieved by either using a ‘transformational reuse’ or a ‘reuse by derivation’. When using a transformational reuse the system identifies the differences between the two problem description of the query case and the retrieved case. The system then alters the solution of the retrieved case with regard to these differences; the system does so by, for example, using replacement rule-based mechanisms.

If the system employs reuse by derivation it analyses the solution of the retrieved cases to implicitly conclude the underlying means that led to this solution.

The system then transforms these means into explicit methods that it then employs to derive a solution for the query case [1]. The adapted solution, forming a new solved case, together with the query problem, is then being validated during the Revise step of the CBR cycle. Based on the design of the CBR system the Revise step can, for example, consist of the solution validation being done by a human expert or it can be done by gathering feedback from the user of the system on the applicability of the suggested solution or it might be automated by using model based validation procedures.

Should the Revise step yield a rejection of the adapted solution the case has to be re-run through the Reuse step to derive a different adapted solution. Should the Reuse step repeatedly fail to adapt the solution into an acceptable, applicable solution the case must be dropped and the system must fall back to the Retrieve step and simply retrieve another ‘next-most similar’ case from its case base. If a revised case yields an applicable solution it should be learned or retained by the system.

3.1 Advantages of using CBR to formalise our specific domain knowledge

CBR is able to ‘*speak the customer’s language*’, allowing for the use of synonyms and missing terms and likely vague terms describing the amount of, for example, a mineral present within a query to our system. In our case this means that fuzzy amount descriptors such as ‘some’ or ‘traces’ can be used to define queries. Additionally CBR is also able to retrieve cases, in *Auric Adviser* descriptions of existing mining operations as well as hydrometallurgical treatments, based on only sparse query data. This is useful while trying to retrieve mining operations only partly specified on sparse prospectation data. Furthermore CBR relies on similarities which are comparatively easy to elicit from human experts within our domain of interest. Finally CBR allows for queries that combine retrieval and filtering in the way of using key attributes as selection criteria for a case before calculating the global case similarity, thus reducing the computational effort of an retrieval. An example for such a pre-selection attribute would be the exclusion of all mining operation cases in which the ore processed does includes clay, as the presence of clay is instantly forbidding a number of chemical treatments of the ore.

A way to amend the described problems of a lack of precisely quantifiable measures and vagueness and only partial prospective data is given by using CBR to also allow for vagueness and a certain amount of ambiguity within the techniques used for retrieving a mining operation best matching the prospective data from a potential mining site and its ore deposit [1,5]. CBR basically is a form of similarity-based retrieval which also allows for a vague query and, given for example a fuzzy case representation, does not require an exact match of a query to produce a result [18,4].

3.2 Knowledge Acquisition

In this section we review available knowledge sources in our domain, we detail on the kind of knowledge we drew from these sources and why we deemed it to be important. We further describe the techniques we used to elicit the knowledge.

We initially identified four sources from which we gathered the following knowledge: Human experts on gold mining, communities of human experts on mining, scientific publications on existing gold ore mining operations, human experts on certain sets of hydrometallurgical treatments and scientific publications on very specific hydrometallurgical treatments. From these sources we gathered the following knowledge:

1. Human experts on gold mining: Necessary attributes and values to describe a gold mining operation
2. Scientific publications on gold ore mining operations: Attribute value ranges and applied ore processing chains
3. Communities of human experts on mining: Similarity measures and similarity measure evaluation
4. Scientific publications on specific hydrometallurgical treatments: Attributes and value ranges to describe treatments and knowledge about their applicability

For our initial knowledge elicitation we used interviews with the human experts. We additionally created questionnaires, combined with similarity measure templates, to be completed by the experts. We had to additionally keep in mind that our acquisition of knowledge must follow certain strict guidelines and suit controlled conditions [3]. To optimise our knowledge elicitation process and techniques we employed an iterative elicitation process. In this iterative process we repeatedly

asked the experts about their feedback on our knowledge elicitation approach, questionnaires and similarity measure templates. The goal of this effort was to allow for input of the experts on how to best enable them to externalise their tacit knowledge. In short we wanted to know if we ‘asked the right questions’. Therefore we asked the experts:

1. Have we asked the right questions?
2. Have we provided the right templates for eliciting the similarity measures (Tables, Taxonomies)?
3. Were our data types and data value ranges correctly set?
4. What input with regard to fundamental knowledge modelling/eliciting did we get from the experts?
5. How would the experts have amended our questions and our way of knowledge gathering?

Based on the feedback, we refined our knowledge elicitation technique and went through a second cycle of knowledge gathering. This iterative knowledge gathering approach is applicable in our work, as the use of *myCBR* Workbench allows us to refine our knowledge model and integrate it in the running *Auric Adviser* software. This is due to the modularisation of the knowledge model component allowed by the use of our *myCBR* SDK.

Our overall knowledge gathering process was again twofold and focused on two areas: First, knowledge about existing gold ore mining operations and the hydrometallurgical process chains used within these mining operations, and , second, empirical knowledge on single hydrometallurgical treatments and their applicability.

The knowledge on existing gold mining operations [8] was used to create an initial episodic knowledge model of mining operations. This knowledge model is now being used within *Auric Adviser* to realise the similarity based retrieval of mining operations based on a query composed from prospective data of a potential mining site. The retrieved most similar gold mining operation’s ore process chain is then recommended for re-use on the potential mining site specified by the prospective data.

The knowledge on specific treatments is to be used to form our second prototypical knowledge model that will be used to realise the similarity based retrieval of single consecutive process steps (treatments), in a very strictly defined and specific context given by a specific ore-constellation and a specific preceding treatment. Thus our second knowledge model will be able

to provide treatment recommendations to solve problems such as the need to ‘reduce sulphides’ for a specific ore/raw product constellation and with regard to a (necessary) preceding treatment already applied to the raw product.

The knowledge necessary for the design of the second knowledge model will be elicited from existing publications on single hydrometallurgical treatments. We assume this source as a valid source of knowledge as when the quality of published work available on single hydrometallurgical treatments was studied [14], the finding was that, apart from some exceptions, hydrometallurgical publications follow a common principle and contain sufficient information needed to describe a treatment’s characteristics.

3.3 Knowledge Formalisation

As already described we aim for a twofold approach of representing the knowledge in our domain. Thus we had to arrange for two different knowledge models serving each of the two aspects of our approach. The first knowledge model which we present in this paper is aimed at holding the knowledge needed to recommend whole ore process chains derived from existing gold mining operations.

Figure 2 shows the approach taken by us to represent and retrieve episodic cases representing existing gold mining operations and their ore refinement process chains. Figure 3 shows the approach which we intend to deploy to represent the case of a preceding and consecutive specific treatment on a specific ore/raw material constellation.

As stated in Section 1, we plan to employ these two knowledge models in a consecutive approach in our final version of the *Auric Adviser*. The first knowledge model is now in use in the *Auric Adviser* in order to retrieve similar existing mining contexts and their ore refinement process chains based on queries composed from prospective data on potential mining sites. Our next development step is now to finalise the second knowledge model to allow for the consecutive automated step-by-step evaluation and, if necessary, re-design of the process chain recommended by the first knowledge model.

As we employed CBR as the reasoning component for the *Auric Adviser* we had to formalise the gathered knowledge into the knowledge representation structure used for CBR, namely the four knowledge containers;

Vocabulary, Cases, Similarity Measures, and Adaption knowledge.

3.4 Initial Knowledge Model

Based on the knowledge gathered from the sources described in Section 3.2 we created our initial knowledge model using the approaches described in Section 3.4. In the following we describe how we formalised the gathered knowledge into the four knowledge containers of any CBR system: vocabulary, cases, similarity measures, and adaptation knowledge [13].

Our vocabulary consists now of 53 attributes, mainly describing the ore and mineralogical aspects of an ore deposit. With regard to the data types used, we used 16 symbolic, 26 floating point, 6 boolean and 5 integer value attributes. The symbolic attributes describe minerals and physical characteristics of minerals and gold particles, such as their distribution in a carrier mineral. Further symbols were elicited to describe the climate and additional contexts a mining operation can be located in, like for example the topography.

We then created a case structure catering for the main aspects of a mining operation, namely: Mineralogical context of the ore, geological context of the ore deposit, environmental context of the mining operation, detailing into: Geographical, topographical, economic and political context of a mining operation. For the initial knowledge model we focused on the mineralogical and geological contexts. The additional environmental contexts are already within our knowledge model but are not yet used as they are intended to be used later on for a more detailed calculation of potential costs of a mining operation. Using this case structure we assembled 25 cases based on mining situations described in literature and from information on such operations provided by experts.

Figure 4 shows a limited excerpt from the case data we generated. Our cases were distinctive mainly with regard to the mineralogical context of the mined ore. Thus we created 5 cases describing refractory arsenopyritic ores, 5 describing free milling gold ores, 2 on silver rich ores, 6 cases on refractory ores containing iron sulphides, 4 on copper rich ores and one each on antimony sulphide rich ores, telluride ore and carbonaceous ore.

To compute the similarity of a query, composed of prospective data, and a case we modelled a series of similarity measures. We had the choice between compar-

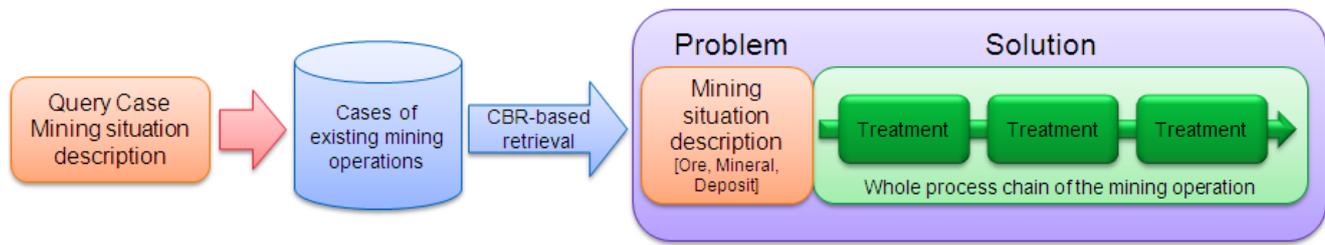


Fig. 2 Case representation of whole ore refinement process chains

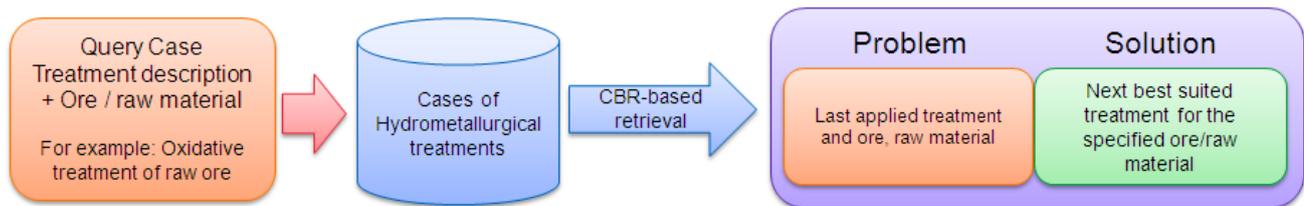


Fig. 3 Case representation of a specific preceding, consecutive pair of treatments

| # | Ore type | Sulfide present | IronSulfide present | ArsenicSulfide present | Free milling gold present | Gold ore grade, g/t | Gold recovery % | Troughput, tpd | Au production oz/yr |
|----|--------------------------|-----------------|---------------------|------------------------|---------------------------|---------------------|-----------------|----------------|---------------------|
| 0 | refractory arsenopyritic | 1.0 | 1.0 | 1.0 | | 8.0 | 88.0 | 1,250 | 100,000 |
| 1 | refractory arsenopyritic | 1.0 | 1.0 | 1.0 | | 7.2 | 91.0 | 1,000 | 75,000 |
| 2 | refractory arsenopyritic | 1.0 | 1.0 | 1.0 | | 7.0 | 85.0 | 465 | 36,000 |
| 3 | refractory arsenopyritic | 1.0 | | 1.0 | | 12.0 | 82.0 | 3,200 | 360,000 |
| 4 | refractory arsenopyritic | 1.0 | 1.0 | 1.0 | | 20.0 | 90.0 | 1,100 | 220,000 |
| 5 | free milling | | | | 1.0 | 8.0 | 98.5 | 8,000 | 730,000 |
| 6 | free milling | | | | 1.0 | 2.4 | 95.0 | 5,000 | 130,000 |
| 7 | free milling | | | | 1.0 | 5.0 | 96.5 | 14,000 | 780,000 |
| 8 | free milling | | | | 1.0 | 1.1 | 80.0 | 41,000 | 400,000 |
| 9 | free milling | | | | 1.0 | 2.0 | 88.0 | 18,000 | 360,000 |
| 10 | silver-rich | | | | | 3.1 | 93.0 | 4,000 | 130,000 |
| 11 | silver-rich | | | | | 70.0 | 95.0 | 360 | 290,000 |
| 12 | refractory iron sulfide | 1.0 | 1.0 | 1.0 | | 5.8 | 91.0 | 1,650 | 100,000 |
| 13 | refractory iron sulfide | 1.0 | 1.0 | | | 3.8 | 89.0 | 5,700 | 220,000 |
| 14 | refractory iron sulfide | 1.0 | 1.0 | | | 4.7 | 90.0 | 2,700 | 130,000 |

Fig. 4 Excerpt from the generated cases

| | AlongCrystalBoundaries | AlongCrystalDefects | IsoidalParticlesInSolution | BetweenGrains | SolidSolution | GrainEnclosedInMineral | Free |
|----------------------------|------------------------|---------------------|----------------------------|---------------|---------------|------------------------|------|
| AlongCrystalBoundaries | 1.0 | 0.8 | 0.5 | 0.9 | 0.7 | 0.2 | 0.6 |
| AlongCrystalDefects | 0.8 | 1.0 | 0.5 | 0.9 | 0.7 | 0.2 | 0.6 |
| IsoidalParticlesInSolution | 0.5 | 0.5 | 1.0 | 0.5 | 0.7 | 0.2 | 0.5 |
| BetweenGrains | 0.9 | 0.9 | 0.5 | 1.0 | 0.8 | 0.2 | 0.6 |
| SolidSolution | 0.7 | 0.7 | 0.7 | 0.8 | 1.0 | 0.2 | 0.7 |
| GrainEnclosedInMineral | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 1.0 | 0.4 |
| Free | 0.6 | 0.6 | 0.5 | 0.6 | 0.7 | 0.4 | 1.0 |
| PartialExposed | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.2 | 0.4 |

Fig. 5 Example of a similarity measure for the gold distribution within an ore

ative tables, taxonomies and integer or floating point functions. For our initial knowledge model we mainly relied on comparative tables.

Our approach includes the idea to model as much of the complex knowledge present in the domain of ore refinement into the similarity measures as possible. This was based on our assumption that the similarity based retrieval approach provided by the use of CBR would allow us to capture and counter most of the vagueness still associated with the selection of the optimal process in the hydrometallurgical treatment of refractory ores domain. For example, we tried to model into the similarity measures such facts as that the ore does not need any more treatment if it contains gold grains greater than 15 micro meters in diameter. Such facts are easy to integrate into the similarity measure and thus are operational (having an effect) in our knowledge model. We deem this capability of the similarity measures to capture and represent such ‘odd’ behaviours of the knowledge model very important. These ‘odd’ facts or bits of knowledge are hard to capture by rules, which may have ultimately kept the IntelliGold approach, which we introduced in section 2, from succeeding on a broad scale [16,17].

For the global similarity measure of our cases we use a weighted sum of the attributes local similarities. We have not yet investigated further on the impact of different weight distributions other than the obvious emphasise of important attributes, such as for example ‘Clay Present’, as the presence of clay forbids a selection of hydrometallurgical treatments.

As we are mainly aiming for case retrieval the need for adaptation knowledge is not yet pressing. We therefore have not formalised any adaptation knowledge. We will however need adaptation knowledge for our second knowledge model which will be deployed to enable the process chain validation and possible re-design. Thus a part of our future work will be to gather and formalise the relevant knowledge to allow for adaptation within our second knowledge model.

With regard to further developing the first knowledge model, or speaking more broadly, enable it to *learn*, the addition of new cases to the knowledge model’s case base will provide this learning ability as within any case based reasoning system.

4 Software Prototype *Auric Adviser*

Using our initial knowledge model we implemented a java-based workflow recommender software, the *Auric Adviser*. *Auric Adviser*’s task is it to retrieve a selection of descriptions of existing gold mining operation best matching the ore and mineralogical context described in a query to the *Auric Adviser*. Once the best matching mining operations are retrieved *Auric Adviser* provides the knowledge about the ore refinement process chains used in these best matching mining operations.

The retrieved case provides the planner of a potential mining operation and subsequent ore refinement process chain with a first draft of what kinds of treatments would be involved in a potential mining on the prospected site. Furthermore the planner gets insights into how to potentially arrange these treatments in a process chain to most efficiently refine the ore at the potential site. Additionally, *Auric Adviser* provides the planner with a possibility to estimate the effort, the costs and some ore refinement constraints involved with the potential mining site.

The *Auric Adviser*’s straight forward user interface allows a planner to specify the data gathered from a prospection into a problem description part and thus compose the query case to be post to the CBR knowledge model (Figure 6). Then the process planner can select the number of retrieved cases she wishes to be displayed. The best retrieved cases are then presented in a tabular field and the process chain description and diagram of the best matching case is displayed in a separate UI element called ‘Solution view’.

We are aware of the complexity of the GUI, offering over 30 elements of input to the planner but we plan to clarify the problem description part of our GUI by breaking it down into the contexts described in our knowledge model, such as *Mineralogical*, *Geological*, *Environmental* etc.

5 Experiments with the *Auric Adviser* and Resulting Refinements to the Initial Knowledge Model

In this section we describe our retrieval experiments with our initial knowledge model. We further detail on the changes we applied to our knowledge model, based on the outcome of the experiments.

The screenshot displays the Auric Adviser software interface, which is used for recommending process chains based on mining and ore context. The interface is divided into several sections:

- Mining Context:** Contains dropdown menus for Mineralogical Context (AntimonySulfide, IronSulfide, AntimonySulfidePresent, CopperSulfidePresent, SulfidePresent) and Deposit Context (Deposit, MainCarrierMineral, Climate, Temp Min C, Temp Max C).
- Ore Context:** Contains dropdown menus for CarbonPresent, SecondaryGangueMat., PrimaryGangueMat., AntimonySulfide, AlongCrystalBoundaries, PrimaryGoldDistr., Secondary G-GrainSize, AlongCrystalDefects, and Primary G-GrainSize. Below these are numerical values for various attributes like % of Austrobitrite, % of Chalcopyrite, % of Digenite, % of LeadIISulfide, % of Pyrite, % of Sphalerite, % of Bornite, % of ClayContent, % FreeMillingGold, % of LeadIISulfide, % of Realgar, % of Stibnite, % of Calaverite, % of Copper, % of GoldContent, % of Marcasite, % of Pyrrhotite, and % of ZincSulfide.
- Process chain diagram:** A flowchart showing the process flow from Cyclone Size Separation and Grinding to Agitated Cyanide Leaching, Thickening, Filtration, and finally to Refining. It includes steps like Atmospheric Carbon Elution, Electrowinning, and Carbon Reactivation.
- Refinement process chain:** A text-based description of the refinement process: "Ore --> crushing, grinding, flotation, roasting, agitated cyanide leaching, filtration, zinc precipitation, smelting --> gold bullion to refinery".
- Case Ranking:** A table showing the best match case. The text indicates: "Auric Adviser found Refiningcase #0 with a similarity of 0.418181818181815 as the best match. The 10 best cases shown in a table:". The table lists various attributes and their values for Refiningcase #0.

At the bottom, there is a "Recommend processchain" button and a "Best Match: 41%" indicator.

Fig. 6 The *Auric Adviser* process chain recommender UI

For example, we performed retrieval experiments with super weighting single attributes, by outweighing their value against all other local similarities of other attributes, to establish the accuracy of a single discriminant attribute. We did so to establish the effective value ranges for the single attribute being analysed in these experiments.

During our first set of case retrieval experiments, we ran 13 test queries on our initial 25 cases, we noticed a set of dominant cases being retrieved disproportional often as the best matches to any given query. The experts advised us to try and add more data to the cases by providing more attribute values but there was still a set of dominant cases. As we re-questioned the domain

experts about this again they advised that we could add even more detail to the cases and refine our selection of the discriminant attributes. By doing so we were able to significantly increase the accuracy and variety of our retrieval results, eliminating the dominant case retrieval.

Based on our experimental data we also noticed that we had to remodel some of the attributes within our knowledge model to represent discriminant Boolean values. These attributes were: *Free milling gold present*, *LeadIISulfide present* and *ZincSulfide present*. We did so to allow for a quicker *exclusion* of cases, based on these boolean attributes.

We also simplified our initial knowledge model by excluding a number of attributes by setting them as non-discriminant in the calculation of the global similarity. We did so to narrow down our knowledge model to the more important attributes and refine the accuracy of these important attributes representations. See figure 7 for the performance data of the 3 versions of our Prototype during the respective retrieval tests.

Starting off from a sub optimal accuracy reported by the experts testing our knowledge model we were able to enhance its performance. After the application of the changes described above we are now getting a more positive feedback from the hydrometallurgical experts. Based on the experiments we are now able to see that our initial knowledge model was over engineered with regard to the number of aspects we tried to cover with it. By reducing these details, focusing on important attributes in the mineralogical context and by refining the value ranges of these attributes we were able to significantly increase the performance of our knowledge model. The results for the queries III and IV indicate to us the need to further refine our knowledge model with regard to the representation of sulphides in it. We deduced from our experiments that we can solve the recommendation problem for the queries III and IV by further adjusting the weight of the sulphide attributes and add additional classifying attributes further detailing the presence or absence of specific kinds of sulphides. We further were able to establish that our knowledge model will also benefit from a finer grained similarity measure modelling of the geological context of an ore deposit, which we plan to implement as future work for the knowledge models further refinement.

6 The second knowledge model

The second knowledge model required for our two fold approach is the knowledge model representing domain knowledge on individual treatments applied to specific types of ore or raw materials whereas a raw material is any kind of partially processed ore before the final extraction of the actual bullion.

As already described the aim of our second knowledge model is it to allow for the step by step validation of a whole process chain recommended by our first knowledge model. Next to the validation aspect our second knowledge model is also intended to recommend better suited, more efficient, treatments based

on more recent research reflected in more recent scientific publications from which the treatment cases are extracted. The idea behind this aspect was that at the time of the creation of the whole process chain that was recommended by our first knowledge model the knowledge on hydrometallurgical treatments had a certain state. If our software recommends the same process two years on in time there might be new insights into certain treatments involved in the process chain. By constantly gathering recent scientific insights from hydrometallurgical publications these new insights into, for example, more effective treatments would then be 'picked up; by our second knowledge model and subsequently would be used in the process chain validation to recommend more effective treatments, based on recent publications on hydrometallurgical treatments. Thus the second knowledge model enables the system to effectively learn the newest specific treatment and apply them to the process chains recommended by the first knowledge model.

6.1 Knowledge sources and Knowledge gathering for the Second Knowledge Model

After surveying the available sources of knowledge for our second knowledge model we chose publications in hydrometallurgical studies of gold extraction, for example scientific papers and industrial reports, as source for cases.

A general principle for a scientific study and for writing a scientific document is repeatability. The study should be describing as detailed as to another researcher should be able to replicate the study and publish results which support or refute the presented theory. When the quality of publication on hydrometallurgical treatments was studied [14], the assumption was that, apart from some exceptions, hydrometallurgical publications follow that principle. It was assumed that most of the publications contain requisite information about treatment characteristics. The outcome of the study was indeed that the majority of publications on hydrometallurgical treatments have a certain, reoccurring, structure. This structure usually consists of the description of an ore or raw material, a treatment or family of treatments applied to it involving a set of substances and parameters of the treatment experiment as well as a description of the results of experimental studies of the treatment describing how effective the treatment was.

| Query ID | #Attributes specified in query | | | Quality of retrieved case** | | |
|----------|--------------------------------|-----------------|-------------|-----------------------------|-------------|-------------|
| | Mineralogical context | Deposit context | Ore context | Prototype 1 | Prototype 2 | Prototype 3 |
| I | 2 | 3 | 2 | 5 | 7 | 7 |
| II | 1 | 3 | 2 | 4 | 7 | 7 |
| III | 1 | 3 | 2 | 1 | 1 | 1 |
| IV | 5 | 2 | 1 | 1 | 1 | 1 |
| V | 0 | 1 | 1 | 7 | 7 | 7 |
| VI | 0 | 0 | 1 | 7 | 7 | 7 |
| VII | 2 | 3 | 5 | 4 | 7 | 10 |
| VIII | 2 | 3 | 4 | 4 | 7 | 7 |
| IX | 5 | 3 | 5 | 10 | 10 | 10 |
| X | 5 | 3 | 5 | 10 | 10 | 10 |
| XI | 2 | 2 | 4 | 4 | 7 | 7 |
| XII | 2 | 3 | 2 | 5 | 5 | 5 |
| XIII | 3 | 3 | 5 | 7 | 7 | 7 |
| | | | sum | 69 | 83 | 86 |
| | | | average | 5.31 | 6.38 | 6.62 |

Fig. 7 Snippet of the performance evaluation of the first three versions of the knowledge model. (* 0 = conflicting, not applicable; 3 = applicable but suboptimal; 5 = applicable, 7 = applicable and well suited; 10 = optimally applicable)

Additionally to the reoccurring structure found it was also found that the quality of the published information about gold extraction varied a lot more than expected [14]. The trends in process description were that gold content and gold recovery were usually well described. The leaching solution and process parameters were not described as detailed. In cases where the treated raw material was a mineral or concentrate, material description was often lacking several details. The mineral type and chemical analysis were described more often in conference proceedings and books than journals.

6.2 Case Structure of the Second Knowledge Model

As mentioned previously we validate the recommended whole process chain in a step to step approach examining the consecutive application of all the individual treatments in the process chain on their subsequent initial ores and resulting raw materials.

We do so to optionally recommend a different treatment if the treatment in the process chain is not matching to a specific ore or raw material expected in our prospective mining side or if there exists knowledge on more effective treatments to replace the treatment applied in the process chain. So we check every treatment

in the process chain if it is the optimal thing to do with the ore or raw material we have at hand. Next to the optimisation aspect we have to perform this validation as we just get a rough recommendation of a process chain from our first knowledge model, based on the mining-context, which uses the rather broad description of a whole mining operation.

Of course if at one place in the process chain we have to recommend another process step based on our more precise knowledge on individual treatments the process chain will break at this point and must be reconsidered, as the replacement of a single process step might result in a different subsequent raw material, for example a chemical solution instead of a foam and thus the following part of the process chain must be redesigned accordingly.

Again, as with our first knowledge model, we chose structural CBR as the technique of choice to represent and retrieve the knowledge on individual hydrometallurgical treatments. Based on the characteristic of an individual hydrometallurgical treatment, frequently described in the literature, we derived the following core attributes for the case structure of our second knowledge model:

1. Ore or raw material properties: A number of parameters describing a raw material

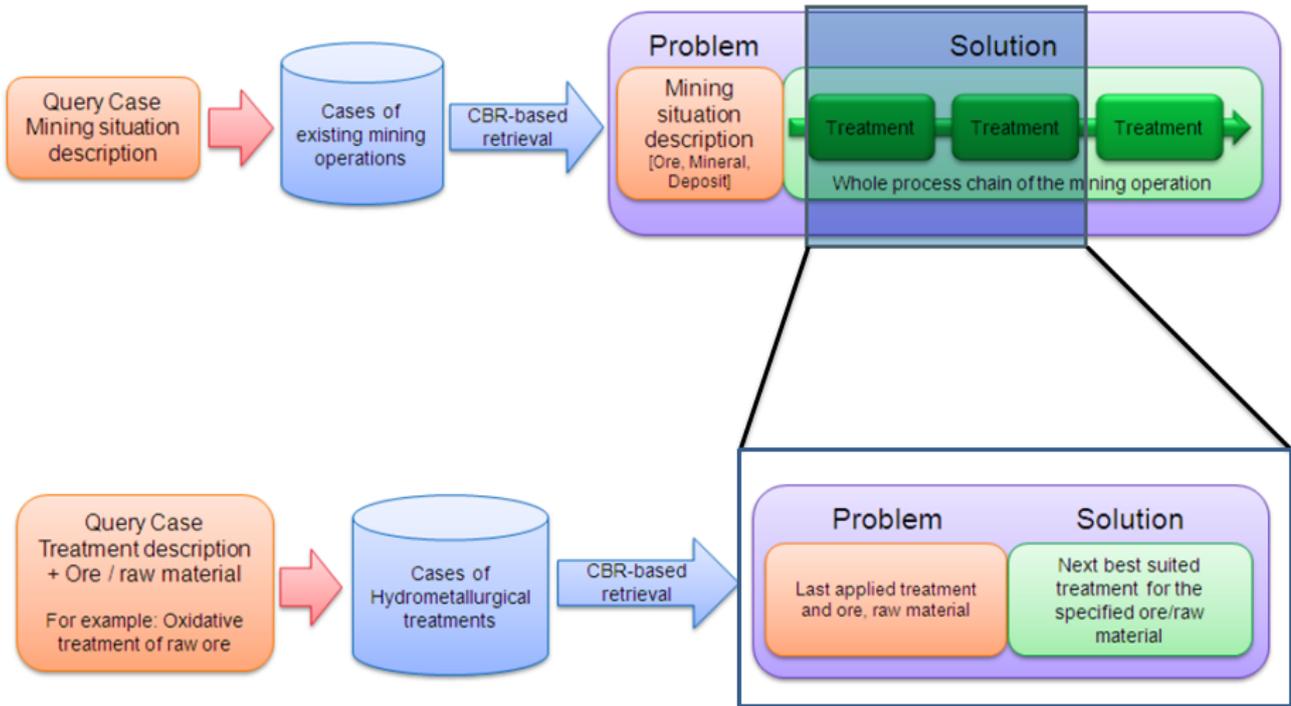


Fig. 8 Step by step evaluation and optimisation of the recommended process chain

2. Treatment: Description of the Treatment that is applied
3. Substances used: Information on the substances involved in the treatment
4. Parameters: Conditional parameters of the treatment, such as temperature, duration or pressure
5. Results: Numerical (per-cent success) or symbolic (“very good”) descriptions of the effectiveness of the treatment

Next to these core attributes that ore case structure is covering we plan to extend the case structure to cover for more additional knowledge on an individual treatment once we have established a sufficiently effective version of our second knowledge model covering the core attributes. Figure 9 provides an overview of the case structure we employ within our second knowledge model.

As stated earlier we aim to derive cases for our second knowledge model on individual treatments from publications on individual hydrometallurgical treatments in a semi-automatic way. To achieve this goal we began our approach with an analysis of the structure of these publications. Starting from surveying the available literature we first derived a hand crafted classification sys-

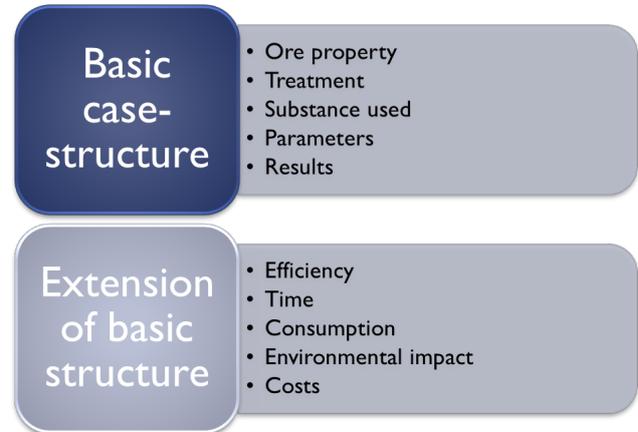


Fig. 9 Case structure describing an individual hydrometallurgical treatment case

tem to identify hydrometallurgical publications of the following types:

1. 1) Clearly structured paper on one specific treatment, following a simple structure like : Ore property-Treatment-Substance-Parameters
2. 2) Papers on a specific family of treatments (like electro-chemical treatments), with a structure like:

- Treatment, Treatment family, Ores, Substances, Parameters”, ’Amount or quantity”, ’Treatment family applicable / not applicable”’
3. 3) Papers on a specific substance or a specific raw material (ore), with a structure like: Ore or Raw Material Preferable, non-preferable treatments or preferable, non-preferable substances
 4. 4) Survey papers on treatments or raw materials or substances, with no definite structure

We aimed to identify these types of publications to be able to automatically classifying papers into these categories with regard to the latter case extraction from these papers. We did so as we deem an individual technique and auxiliary data necessary for each category of publications to extract cases from these categories.

For the semi-automatic extraction of cases from papers that are of the Grades 1 or 2 we are currently employing a customised ANNIE² information extraction application, which is part of the GATE natural language processing framework developed by the University of Sheffield. The purpose of this application is to classify literature on treatments into the described specific categories of treatments. To being able to do this the ANNIE Application is designed to annotate certain keywords from categories as: Substance names, ore or raw materials, parameters, keywords from result descriptions. To enable this automatic annotation we created a set of auxiliary data that is used by the ANNIE application to annotate terms from the categories described above. Such auxiliary data is for example a set of Gazetteers for treatment names, Substance names, Ore specifications, Parameters and terms often found in descriptions of results. Additionally to these Gazetteers we created a set of Jape rules, Jape being part of the GATE framework, to identify terms from these categories that are not covered by the Gazetteers as well as to perform basic stemming.

7 Discussion

In this section we will analyse the gathered data from our usage experiments on the *Auric Adviser*. We evaluate the usability and performance/accuracy of our knowledge model and subsequently our *Auric Adviser* software. We do so based mainly on initial feedback from experts in the field of gold mining and hydrometallurgy

as our prototype’s knowledge models are not yet scaled to a scale that would allow for the gathering and quantitative analysis of usage data. We therefore rely on the qualitative analysis of feedback we gathered from experts that used our prototype.

7.1 Knowledge Modelling: Feedback on *myCBR* Workbench

After presenting the prototype software to selected experts we were able to obtain the following first feedback from the domain experts. The experts were generally satisfied with the accuracy of the recommendations provided by our prototype. The experts further were particularly pleased with the ability of our prototype to provide very suitable workflow recommendations on very sparse queries entered into the prototype.

The graphical user interface of *myCBR* Workbench with which we modelled our domain knowledge, was seen as logical and easy to follow. Adding new concepts and attributes to our knowledge model, was deemed simple and effective by the experts. Additionally the experts noted that it was very useful that the concepts and attributes are shown in alphabetical order in both, knowledge modelling view and case base view.

However the experts also noted that when re-examining the knowledge model and when adding new cases, it would be more practical if the user could organise concepts and attributes in a logical order. This custom ordering of a case’s attribute would also ease the process of feedback gathering from domain experts as they could prioritise attributes and thus review them faster. By allowing for a custom sorting of the case attributes the user could focus on adding correct values into the cases, instead of concentrating on the rigid order of attributes provided by *myCBR*. Additionally some of the experts suggested using the same interface for case adding as it is used now for retrieval testing within *myCBR*.

Feedback on the User Interface of the Prototype: The experts suggested that a user should be able to choose if the attributes ‘ore throughput rate’ (tons per day) and ‘gold recovery’ are included in the similarity calculation or not.

The assumption behind this was that the experts deemed the volume of ore and its gold content two very important factors with great impact on the design of a process chain. The experts recommended using a fixed

² <http://gate.ac.uk>

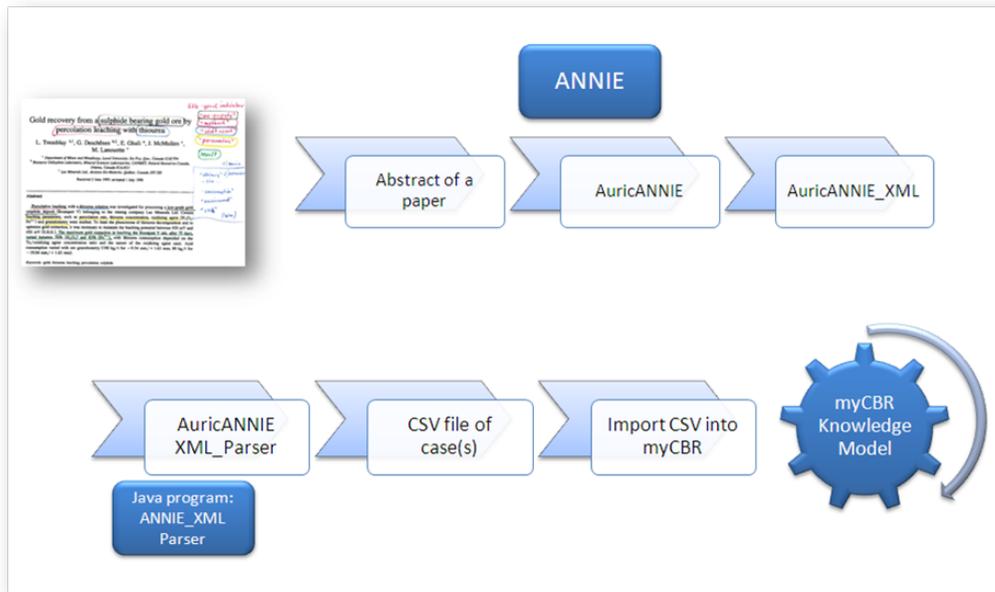


Fig. 10 Semi-automatic case extraction process for the second knowledge model

set of terms for the display of the description of the process chain within the solution part of our cases.

With regard to case ranking the experts advised that we could simplify the case representation in our initial knowledge model. The experts recommended using a table display for the case ranking. Furthermore the readability of our solution display could be further enhanced by relying on colour coding of the local similarities or the use of other visual aids.

An additional feature suggested by the experts for our *Auric Adviser* prototype was the ability to save a query and its resulting retrieval result for later reuse and export, for example into an Excel table. We plan to integrate this useful feature as one of the next development steps within our software.

We are currently in contact with the hydrometallurgical experts to gather feedback on the first iteration of our second knowledge model. We do so by, again, the use of questionnaires, asking the experts about the applicability of recommended treatments on a given ore or raw material as well as for alternative suggestions by the expert. We do so to test the accuracy of our second knowledge model, by retrieval testing and feedback from experts, as well as to establish if our knowledge model still "misses" opportunities to recommend bet-

ter treatments, where this is possible within the recommended whole process chains.

As stated in section 2, the only system implemented yet, being directly comparable to our prototype, was the IntelliGold system created by Torres et al. [16,17]. The intention of the IntelliGold system was focused on recommendations on the costs of a potential mining site, whereas our approach is focused on the recommendation of workflows for a potential mining site. It is also noteworthy that the IntelliGold system followed a wider approach, trying to model more information on a mining site, whereas our approach strictly focusses, yet, on the hydrometallurgical workflow to be employed in a potential mining site. However in this context of recommending hydrometallurgical workflows, our case-based reasoning system appears to be more versatile than the rule based approach employed in the IntelliGold approach. The reason for this was identified as the capability of the case-based approach to better handle sparse data than a rule based system would be able to. Additionally the similarity measures employed within our case-based reasoning approach proved to be more capable to model the 'odd' facts in the domain of hydrometallurgical processing than a purely rule based system would have been.

8 Summary and Outlook

We presented our twofold approach to eliciting and formalising knowledge in the domain of hydrometallurgical processing of gold ore. We did so based on the relevant work in the fields our research touches, namely structural CBR, hydrometallurgy and workflow recommendation.

We demonstrated our processes of formalising the captured knowledge and detailed the resulting first knowledge model and its use in the *Auric Adviser* workflow recommender software. We established its usability and the quality of its recommendations as well as the accuracy and performance of our knowledge model in a set of first experiments.

Based on initial feedback we have shown that our first knowledge model was slightly over engineered and lacked in the quality of its recommendations. Fortunately, based on the detailed feedback we gathered from experts in the gold ore treatment domain, we were able to significantly increase the accuracy and performance of our knowledge model. We provided proof for this enhancement, by detailing on the knowledge model refinements, within the evaluation section of this paper.

We also demonstrated the composing of the second knowledge model for recommending single treatments, based on ‘lessons learned’ from the development of the first knowledge model. Parallel to this we will further refine our first knowledge model, used for retrieving process chains based on prospective data.

All in all we are now confident with the on-going development of our CBR knowledge models and are now working on gathering and formalising further knowledge from the domain. We plan to employ this knowledge in our future work to further enrich our knowledge models. When both knowledge models are finalised we will combine them into a new and complete version of the *Auric Adviser*.

Acknowledgements This work was supported in part by the LOWGRADE project of the ELEMET research program funded by FIMECC Oy. The financial support of TEKES and Outotec Oyj is gratefully acknowledged as well as the financial support of Technology Industries of Finland Centennial Foundation Fund for the Association of Finnish Steel and Metal Producers.

References

1. Aamodt, A., Plaza, E.: Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI Communications* **1**(7) (1994)
2. Arcos, J., Grachten, M., de Mántaras, R.: Extracting performers behaviors to annotate cases in a cbr system for musical tempo transformations. *Case-Based Reasoning Research and Development* pp. 1066–1066 (2003)
3. Darke, G.: Assessment of timbre using verbal attributes. In: *Conference on Interdisciplinary Musicology*. Montreal, Quebec (2005)
4. Hirota, K., Yoshino, H., Xu, M., Zhu, Y.: An application of fuzzy theory to the case-based reasoning of the cigs. *Journal of Advanced Computational Intelligence* Vol **1**(2) (1997)
5. Lenz, M., Bartsch-Spörl, B., Burkhard, H.D., Wess, S. (eds.): *Case-Based Reasoning Technology: From Foundations to Applications*, *Lecture Notes in Artificial Intelligence*, vol. LNAI 1400. Springer-Verlag, Berlin (1998)
6. Lopez De Mantaras, R., McSherry, D., Bridge, D., Leake, D., Smyth, B., Craw, S., Faltings, B., Maher, M.L., Cox, M.T., Forbus, K., et al.: Retrieval, reuse, revision and retention in case-based reasoning. *The Knowledge Engineering Review* **20**(03), 215–240 (2005)
7. Madhusudan, T., Zhao, J., Marshall, B.: A case-based reasoning framework for workflow model management. *Data & Knowledge Engineering* **50**(1), 87–115 (2004)
8. Marsden, J., House, I.: *The chemistry of gold extraction*. Society for Mining, Metallurgy, and Exploration (2006)
9. Minor, M., Bergmann, R., Görg, S., Walter, K.: Towards case-based adaptation of workflows. *Case-Based Reasoning. Research and Development* pp. 421–435 (2010)
10. Minor, M., Tartakovski, A., Bergmann, R.: Representation and structure-based similarity assessment for agile workflows. *CBR Research and Development* pp. 224–238 (2007)
11. Pajula, E., et al.: *Studies on computer aided process and equipment design in process industry* (2006)
12. Plaza, E., Arcos, J.: Constructive adaptation. *Advances in Case-Based Reasoning* pp. 306–320 (2002)
13. Richter, M.M.: Introduction. In: M. Lenz, B. Bartsch-Spörl, H.D. Burkhard, S. Wess (eds.) *Case-Based Reasoning Technology – From Foundations to Applications*, LNAI 1400. Springer-Verlag, Berlin (1998)
14. Rintala, L., Aromaa, J., Forsen, O.: Use of published data in the development of hydrometallurgical flow sheet for gold using decision-support tools. In: *Proceedings of the International Mineral Processing Congress, IMPC 2012*. CSIR (2012)
15. Sauer, C.S., Roth-Berghofer, T., Auricchio, N., Proctor, S.: Recommending audio mixing workflows. In: *Proceedings of the 21st International Conference on Case-Based Reasoning (ICCBR 2013)*. Springer (2013)
16. Torres, V., Chaves, A., Meech, J.: Intelligold-a fuzzy expert system for gold plant process design. In: *Fuzzy Information Processing Society, 1999. NAFIPS. 18th International Conference of the North American*, pp. 899–904. IEEE (1999)
17. Torres, V.M., Chaves, A.P., Meech, J.A.: Intelligold-an expert system for gold plant process design. *Cybernetics & Systems* **31**(5), 591–610 (2000)

18. Watson, I.: Case-based reasoning is a methodology not a technology. *Knowledge-Based Systems* **12**(5), 303–308 (1999)