The Support of Higher-level Cognition in the Context of Ill-structured Process Knowledge

Ronald Grau R.R.GRAU@KINGSTON.AC.UK
Computing and Information Systems, Kingston University London, Penrhyn Road, KT1 2EE, UK

Peter Cheng P.C.H.CHENG@SUSSEX.AC.UK
Cognitive and Language Processing Systems, University of Sussex, Brighton, BN1 9QJ, UK

Abstract

How can knowledge acquisition, modeling, problem solving, and discovery be facilitated in domains where systems involve complex, heterogeneous, and multi-scale processes? This paper presents the conceptual developments of a research program that aims to support higher-level cognition in such demanding contexts. These process domains present a major challenge for human cognition because the related knowledge is ill-structured and spans different scientific areas that employ a wide spectrum of paradigms and models. Further, diverse notations are required to encode the knowledge. Our approach describes a problem space for performing knowledge acquisition in such domains and is set in a wider context of modeling and discovery. We have developed a framework of structured representations, methods, and heuristics, which are utilized to elicit and encode ill-structured process knowledge in a largely diagrammatic knowledge-based system, and to provide human-level interfaces for an interactive software program, designed to aid humans in the specification, inspection, and exploration of ill-structured process knowledge. Specially designed diagrams are fundamental to the approach because they provide a neutral notation to integrate many aspects of heterogeneous knowledge, whilst exploiting cognitive and representational properties that provide advantages for human understanding and reasoning. Evidence for the utility of the approach was found as a result of an initial knowledge acquisition case study in a representative industrial domain that involves combinations of physical, chemical, and biological processes.

1. Introduction

Ill-structured problems have been characterized as those where the related knowledge is difficult to structure and decompose into manageable parts for closer examination and study (Simon, 1973). These kinds of problems resist the application of narrowly defined concepts, operators, sets of rules, or principles for finding a solution, and often have multiple solutions and solution paths (Jonassen, 1997). This paper addresses their occurrence in the context of the human understanding and exploration of complex processes where the related knowledge is intimately intertwined across many levels of abstraction, and where explanations require several perspectives to be considered. In particular, domains that involve heterogeneous, mutually interacting components and processes which exhibit effects dynamically and on multiple temporal and spatial scales present a tremendous challenge for human cognition. Performing systematic knowledge operations in these domains is a holistic, ill-structured problem, exacerbated as particular concepts, paradigms, models, and notations inherent to different scientific or technological areas need to be considered in combination.

Such processes are ubiquitous in both industrial and scientific domains. For example, certain manufacturing domains like Pharmaceutics, Food Processing, or Materials Synthesis engage multistage processing to combine diverse materials into products, where both the materials used and also the methods and procedures involved in their transformation directly influence the course of processes which have an impact on the final structure, functions, and properties of a product. Although some related scientific knowledge exists about the processes involved, it appears distributed across different disciplines and is difficult to represent and examine in combination. As a result, interactions between heterogeneous processes are often not fully understood such that the design of products relies on the tacit knowledge from experienced practitioners and researchers, and the application of weak trial-and-error methods in order to make up for uncertain, vague, or unreliable knowledge about the domain. In science, ill-structured process knowledge presents a common problem for disciplines that need to consider data and models from different areas, especially at the forefront of scientific discovery. Examples include research in areas like Systems Biology, Climatology, Geology, or Astronomy. As these fields are increasingly drawing together scientific data from multiple sources – inheriting their related concepts, paradigms, models, and notations – a need arises for novel methods and tools that help integrate and interpret these data, and discover new knowledge.

Despite the apparent importance of ill-structured process knowledge, there have been only few attempts to develop new conceptual approaches, paradigms, or computational tools to support high-level cognition in such domains. While there has been some research in areas such as Cognitive Science and Artificial Intelligence, Process Modeling, Knowledge Engineering, Information Visualization, or Computational Scientific Discovery, which has addressed related aspects in isolation, none of these areas present an overall solution. A major reason for this common but unfortunate state of affairs lies in the lack of adequate characterizations of the diverse and complex knowledge that would be needed to underpin any systematic approach to the modeling and exploration of the related processes. The notion of *complex, multi-dynamic (CMD) processes* (Grau 2009) presents an attempt to characterize the ill-structured nature of these processes as classes of knowledge, and to capture the various abstractions and perspectives that exist in the context of structural, behavioral, and functional aspects of complexity, as well as some of the dynamic interrelations between these aspects. For example, consider systems that involve structural dynamics, where the conceptual and the physical structure diverge over time as materials and their subcomponents develop multiple physical instances which become spatially distributed over different parts of the system; and where new materials may be created, existing materials destroyed, or properties be modified in the course of processes. CMD processes also capture notions of behavioral and functional dynamics, where the activation conditions and the nature of the processes associated with particular substances and functional outputs depend on environmental conditions, the properties of existing components, and the performance of other processes. These may compete for the same resources, and possibly exhibit magnifying or opposing effects on different properties of components, other processes, or the system as a whole.

The general approach described in this paper aims to achieve some first steps in providing a comprehensive conceptual, representational, and cognitive foundation for the development of methods and tools that assist humans in the acquisition, modeling and discovery of ill-structured process knowledge. As such, its purpose is not to simulate human problem-solving expertise or system behavior but to provide knowledge-level interfaces to computational tools which exploit specially designed representations, methods, and heuristics that facilitate systematic knowledge operations in the context of suitable problem spaces. The developments converge in an interactive, computational tool kit for the support of higher-level cognition. CMD SUITE, a first prototype, was developed which captures specified knowledge in a largely diagrammatic, knowledge-based system. In the following section 2, the approach and its current implementation will be described. Section 3 will then discuss the scope of the research and current limitations; extensions and implications for related fields; as well as future applications.

2. Approach

In this section, we introduce in general terms the key ideas underpinning the theoretical framework and some of the methods and tools developed for the sample task of knowledge acquisition. Further, we provide an illustration of some of the major functions provided by the software workbench to support this task. Finally we briefly summarize the findings of an industrial case study, carried out as an initial evaluation of the approach (Grau 2009).

2.1 Central Ideas and Aims

A core objective of the current approach is to provide a framework for addressing two initial major problems for the capture of ill-structured process knowledge. Firstly, the framework requires *a system of structured representations* to allow the specification and encoding of heterogeneous knowledge in a systematic fashion. Secondly, the framework needs to define *methods and heuristics* for users to operate within the problem spaces which correspond to different high-level cognitive tasks, specifically the acquisition, modeling and discovery of knowledge about a particular process system. We contend that all of these need to be carried out in concert for any single task to be successful. In this general context, a number of relationships between lower-level knowledge operations and high-level tasks can be identified (*Figure 1*).

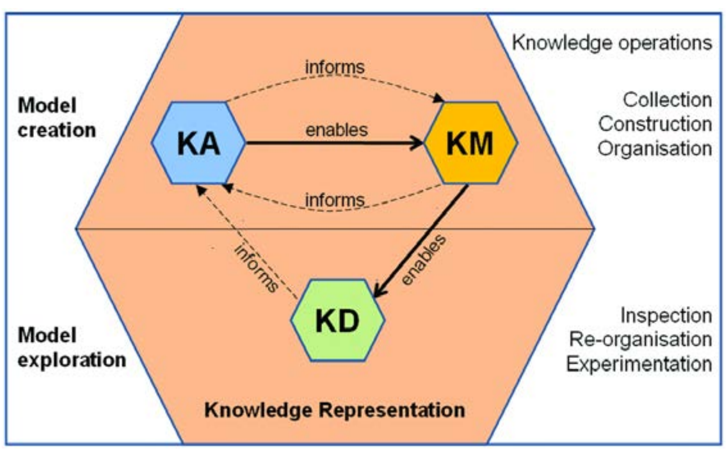

Figure 1: Relations Between High-level Cognitive Tasks in Domains with Ill-structured Process Knowledge (Source: Grau 2009)

It was found that modeling the inherent characteristics of ill-structured process knowledge requires a compositional approach (Falkenhainer and Forbus, 1991), as the relevant knowledge is fragmented and largely dependent on specific contexts of interacting components, processes, and related outcomes (Grau, 2009). As there are only few specifications of components and processes available initially to start knowledge acquisition activities with, a generic model paradigm (or meta-model) needs to be developed, reflecting the overall conceptual structure of the process domain. This should be based on a system of integrated knowledge representations of well-defined facts and relations about the domain (Represented by the large hexagonal structure in *Figure 1*), and is essential to apply a working structure to the ill-structured problem of defining useful problem spaces and organizing the lower-level operations in these spaces (cf. Simon 1973).[1] In the following, let us look at the specific interrelations indicated in *Figure 1*:

- KA-KM:
  Knowledge acquisition (KA) and modeling (KM) are deeply related, mutually dependent tasks, in that any knowledge acquired needs to be integrated into a joint model, whereas additional cues and constraints resulting from a growing compositional model will further inform the knowledge acquisition process: As additional facts and relations are entered, organized, and interrelated, these will provide valuable contexts for the specification of additional process fragments.[2]
- KM-KD:
  A model of sufficient size will eventually enable operations for its inspection, reorganization, as well as for further experimentation, i.e. the exploration of the constructed model. This may lead to the discovery (KD) of previously unknown facts and relations about the domain.
- KD-KA:
  Newly discovered knowledge may provide insights about structural, behavioral or functional changes within the system of processes which were not considered previously. As such, discovery could inform further knowledge acquisition (KA) activities by providing new contexts for the specification of process fragments, and the subsequent revision of those fragments that already exist in the model.

Whereas the original idea of problem spaces in computational scientific discovery is that of a search of system states, determined by the application of specific operators, and governed by heuristics (Newell and Simon, 1972; Klahr & Dunbar, 1988), such approaches cannot directly be applied to bodies of ill-structured process knowledge. This results from the fact that the states of systems involving complex multi-dynamic (CMD) processes cannot be defined in full initially, as they depend on multiple contexts of components and processes, as well as the outcome of their dynamic interactions. Such knowledge is usually not available at the beginning of modeling activities, and would only fully emerge once all knowledge has been specified about such a system. The cognitive operations supported by our approach eventually also constitute an attempt to find interesting problem spaces which may be a basis for subsequent computational

[1] A video explaining this problem further, as well as demonstrations of our related software implementation are available at the URL `http://tinyurl.com/q7b6fup` ,which links to a YouTube playlist.

[2] For instance, consider a partial process describing some chemical reaction, where one reaction outcome may be a particular material involved in another, not yet specified, process fragment.

exploration. In this fashion, finding the conceptual definitions and relations of the components and mechanisms which govern the structure, behavior, and function of an overall system of processes directly corresponds to the initial stages of a discovery process (Langley 1998): The forming of reliable taxonomies, and the gaining of a qualitative understanding about the workings of a domain or phenomenon.

In its first installment, our approach aims to develop suitable *representations* for ill-structured process knowledge, as well as *methods and heuristics* to define a problem space for performing initial *knowledge acquisition* activities. Specifically, this requires the:

1) Identification of the conceptual dimensions (i.e., orthogonal sets of core concepts) which govern the available knowledge about the domain.
2) Development of a generic model paradigm that reflects the conceptual structure of the domain, and that allows the systematic specification of processes, while integrating the individual notations which are commonly employed in their description.
3) Development of representational schemes to codify the various kinds of knowledge corresponding to the conceptual dimensions, and the relations between them.
4) Definition of a problem space for knowledge acquisition, underpinned by specific representations, as well as related methods and heuristics for the capturing of knowledge from human experts.

It would not have been possible to develop this approach in an abstract fashion, i.e. without knowledge input and expert guidance from a representative application domain. The work was substantially informed by the domain of food processing, specifically industrial baking, and the related intersecting physical, chemical, and biological processes in this domain.

2.2 Representational System

The framework employs and extends methods of Representational Epistemic Design (e.g., Cheng & Barone, 2007) for creating representational systems. In contrast to notations based on algebra or logic, for example, specially designed diagrammatic representations can support the integration of knowledge that appears both qualitative and quantitative; the externalization of tacit knowledge in a form that supports reflection and insight; the expression of knowledge in terms of relations and constraints; and the specification and interrelation of alternative contexts for the same or alternative aspects of given knowledge (e.g., Cheng & Simon, 1995; Cheng, 1996; Cheng, Cupit & Shadbolt, 2001; Vicente & Rasmussen, 1992).

2.2.1 Conceptual Dimensions

The design of representations is substantially informed by the conceptual dimensions that govern a domain. Those identified for industrial baking are shown in *Table 1*, and ordered according to those dimensions representing ontological individuals (Type I), and those which may require multiple conceptual dimensions to be combined and interpreted (Type II) in their specification. The conceptual dimensions were devised in collaboration with domain experts and underpin the knowledge in the domain in terms of its explicit conceptualizations as well as the implicit assumptions about different related sub-domains of knowledge. The third column shows where the knowledge associated with a particular dimension is represented or used in CMD SUITE.

Table 1. Conceptual Dimensions identified for the domain of Industrial Baking.

Conceptual Dimension	**Examples of related knowledge**[3]	**Corresponding CMD SUITE tools or representations**[4]
Type I		
Taxonomy	Bakery products Classes of ingredients Process-relevant materials Intermediate products Manufacturing hardware Environmental conditions	Library Diagram, General Knowledge Base tool
Physical Space	Spatial location of materials Physical proximity of materials *Mass flow*	Partitioning Diagram
Processing Time	Activities described in a bakery recipe Duration and sequencing of processes *Speed of interaction* *Simultaneous and subsequent processes*	Context Information Box, Temporal Scope Diagram
Properties	Physical properties Chemical properties Abstract properties Structural, behavioral and functional effects	Compound Diagram, Scenario Diagram, Model Component Editor; Different perspectives, and flexible levels of abstraction
Property Scales	Property unit scales Possible value ranges of properties Vague and incomplete knowledge about behavior	Scenario Diagram (Quantity spaces), Compound Diagram, Model Component Editor
Type II	**Examples of related knowledge / dimensions**	
Structure	Taxonomy (I) Physical Space (I)	Compound Diagram Partitioning Diagram
Behavior	Taxonomy (I) Physical Space (I) Processing Time (I) Changing properties *Coinciding processes or related events* *Scientific laws* *Causal relations*	Library + Partitioning Diagram; Different resolutions in the Temporal Scope Diagram; Quantity spaces for properties; Line graphs denoting changes to property values; Event marks denoting sequencing and time
Function	Taxonomy (I) Physical Space (I) Processing Time (I) Structure (II) Behavior (II) *Production goals* *Means-ends relationships*	Structural configuration displayed in the compound diagram; Abstract properties in the scenario diagram - different levels of abstraction and granularity; Finite processing time

[3] Examples shown in italic face are not yet represented in the current set of tools. However, these have been considered in the original design to be implemented in future extensions of the framework and the tools.
[4] See sections 2.2.3 and 2.3 for details.

2.2.2 Generic Model Paradigm

The definition of a generic model paradigm allows for the construction of a bare-bone knowledge structure to organize knowledge acquisition and modeling tasks for the domain and to integrate heterogeneous knowledge. The paradigm for CMD processes is based on the following basic principles:

- Many mechanisms that relate to structural, behavioral, and functional aspects of processes can be made explicit by adopting a mechanistic approach to their description and explanation, considering the physical changes and effects that can be observed in the system by scientific means (Bechtel and Richardson 1993).
- As described in the context of near-decomposable systems (Simon 1962), the physical proximity between identified materials can be an indicator for the probability of strong (i.e., observable within a limited time frame) interactions that involve these materials.
- The external application of energy to a system of processes can be an indicator for the initiation of new processes, or a change of properties affecting existing processes.

On the basis of these principles, a central part of the approach is the initial hierarchical and temporal decomposition of a CMD process in order to derive a first set of process contexts for knowledge acquisition – or in other words, to apply a structure to this otherwise ill-structured task. In industrial baking, the algorithm-like characteristics of *product recipes* can be exploited to decompose and extract such knowledge objects. Generally, this concerns any planned material inputs and intermediate output components and the operations upon them, and constraints in terms of relevant property states of components and temporal restrictions for operations. The aim of this initial recipe decomposition is to create a time ordered description of substances and activities involved in the making of a product, and to make explicit important transition points where there is a high probability of strong interactions occurring (e.g., when there are materials added to the system, energy applied, or a change to environmental conditions). Although defining such a time-ordered segmentation is based solely on assumptions about the occurrence of processes under the paradigm, it is important to perform this task because there is usually no initial knowledge available about the exact start and end time of single interactions in the system. However, expert statements about partial processes are often based on contextual knowledge about the composition of matter and changes of matter and energy. A suitable and easily accessible representation of this knowledge (such as the *Compound Diagram* described in the next section) can make such contexts instantly accessible for domain experts in order to constrain the size of interaction contexts, and act as an aid for determining approximate time slots for the correct specification and placement of lower-level knowledge fragments in the model.

2.2.3 Representations

While the framework features a whole range of different tools and related representations, we present here but three particular diagrams to illustrate their utility.

The *Compound Diagram* shown in *Figure 2* (right) has been constructed from the knowledge objects extracted from a product recipe of a pound cake (*Figure 2*, left) – a simple cake made of equal amounts of sugar, flour, margarine and whole egg. The circles on the left are just for illustration and not part of the diagram. They show different hierarchical temporal segments identified in the recipe, and the particular changes and activities associated to these temporal segments. Each colored rectangle in the Compound Diagram represents the availability of a particular substance or ingredient at a specific processing step. In their computational, interactive form, compound diagrams can be generated for different structural abstraction levels and represent the structural configuration of a product along a sequence of temporal periods, which constrains the size and amount of initial contexts available for the specification of process fragments. Compound diagrams also distinguish initial ingredients and intermediate products. When intermediate products are created during a process, these would be added to the compound diagram as newly created components and placed within the time slot in which they appear. This is a powerful and flexible way to capture structural dynamics diagrammatically. Configurations of elements which are in close physical proximity can be selected in order to define a particular process context that may combine different substances and span over one or more recipe steps.

Initially, Compound Diagrams are useful for providing starting points for knowledge acquisition, in particular for the specification of lower-level knowledge about the processes that occur within a specific temporal context, involving the components selected by the expert user.

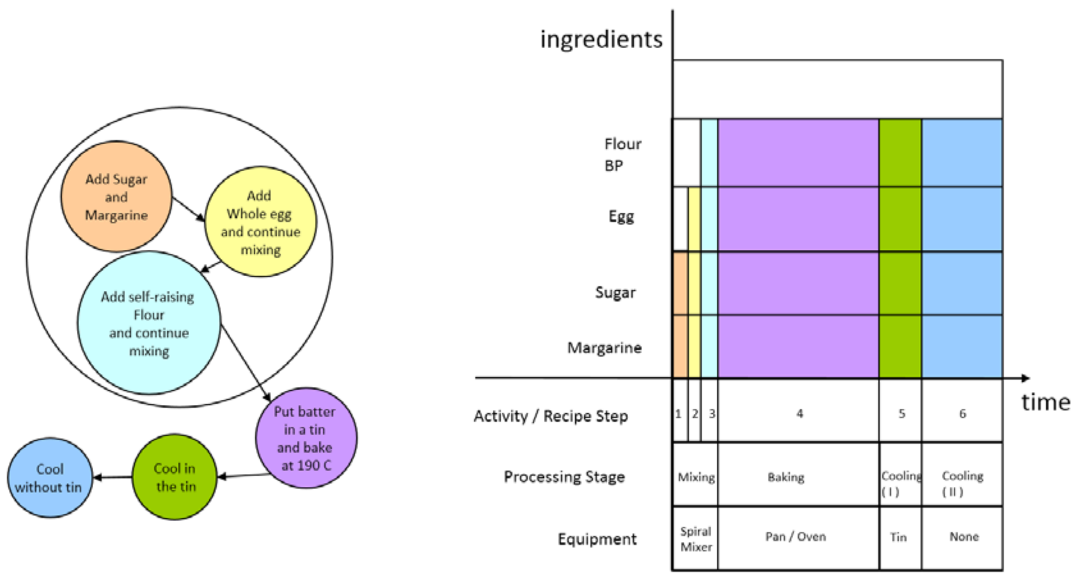

Figure 2. Representing a Decomposed Recipe in a Compound Diagram (Illustration)

Scenario diagrams (*Figure 3*) are representations of process fragments, which combine different materials and allow the recording of changes to their properties within a specified temporal period and in relation to a succession of discrete process events. The sequencing of events is marked by the blue points in the Temporal Scope tool (bottom). Event marks are introduced when the expert makes changes to property values of structural components. Each property scale is represented by a quantity space – an idea adapted from qualitative modeling (e.g., Forbus, 1984) for denoting changes to properties qualitatively within meaningful boundaries. Quantity spaces are considered a flexible notation to represent knowledge at different resolutions, matching the varying completeness and precision of available expert knowledge.

The diagram in *Figure 3* shows a number of parallel interactions within a fragment such as the effects of an ongoing sucrose solution process that leads to an increase in the saturation of "Water", and a decrease in the crystalline particle size of "Sugar". There is also a general decrease indicated to the amount of solid fat remaining in the material "Margarine", which could be further related to other process fragments concerned with the temperature or plasticity of this material. There is also a new ingredient called "Incorporated Air" which was introduced into the system, with its volume increasing to a certain maximum, and falling afterwards. The temporal event mark that is related to the peak was given a quantitative value (35s) by the expert that indicates a potentially interesting state for the inspection of other component properties within the same, or other related fragments. Individual property values can be annotated using labels which

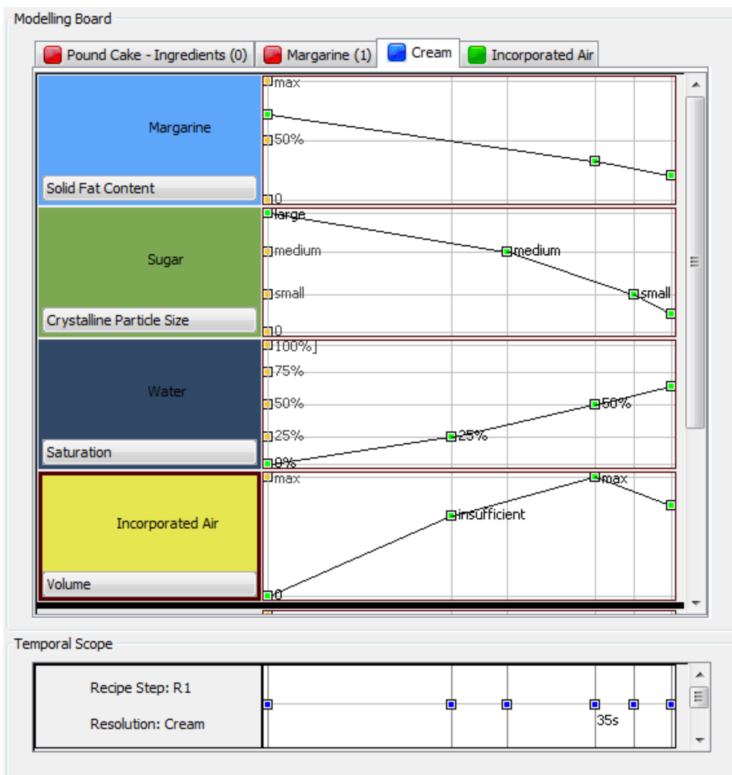

Figure 3. A Scenario Diagram (Partial CMD SUITE Screen Snapshot)

show the corresponding quantity space value, or individual notes. The components stacked vertically in a scenario diagram can be distinguished into two types: *Non-material* components can be introduced to denote property changes at other levels of abstraction (e.g., the product compound as whole could be added to the fragment as a component to say something about a corresponding change to its overall volume, for example). On the other hand, *material* components add to the physical partition defined by the fragment, and so to a new material defined at a higher level of abstraction. This can then be used for knowledge specification in other process fragments. For instance, in *Figure 3*, the combined partitions of fat, sugar and air specify another meaningful material – cream. The scenario diagram also details the environment of the components in the system by integrating the related materials (e.g. Environmental Air) and relevant properties (e.g., temperature), as specified in a decomposed recipe.[5]

Partition diagrams (*Figure 4*) represent the physical boundary configurations of materials, which may be subject to change in the course of processes. When generating a partition diagram, the software determines the material component instances that are available at a selected point in time during processing, with each instance of the diagram corresponding to a specific temporal event mark. The diagram displays the physical partitions of components as rectangular shapes. These can then be resized and rearranged by the user to define a particular spatial configuration.

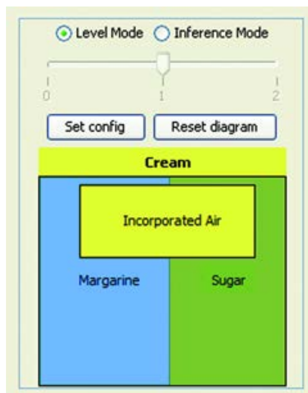

Figure 4. A partition diagram example (Partial CMD SUITE screen snapshot)

After an arrangement has been confirmed, the shapes are analyzed by an algorithm that determines which rectangles overlap (physical contact), which are enclosed by others (isolated from other components), and which have been placed at the border of the diagram (possible interactions components in other process fragments). This information is then stored in the knowledge base including 1) the coordinates, width, and height of shapes for the purpose of recreating the diagram instance later, if needed, and 2) a boundary configuration list denoting those components in direct contact with others, and those which are isolated from other components at this point in time. This data will later be exploited to determine sets of components from this and other process fragments which are likely to interact, and point towards new prospective interaction contexts which should be inspected and validated by the human user.

The three diagrams presented in this section are part of an integrated representational system that facilitates the capturing of different aspects of structural, behavioral, and functional knowledge about the processes in the system through direct, human-driven knowledge input. As the corresponding tools are interlinked, interactive feedback is given to a user once particular modeling decisions made in one tool have an impact on another, which is useful for providing an instant overview of the changes in the system and guide the next modeling activities.[6]

[5] The environmental components are located below the bold, black bar and not directly visible in *Figure 3*.
[6] For instance, a component created by a chemical reaction specified in the scenario diagram would be reflected in the compound diagram, and be available as a possible context for another process scenario.

2.3 CMD SUITE

In this section we will briefly describe the software workbench developed and, at the same time, illustrate some of the major operations and heuristics facilitated for the higher-level task of knowledge acquisition.

2.3.1 Brief Overview of the Workbench

A screen snapshot of the workbench is shown in *Figure 5*. The main areas comprise a library tool (left), displaying the materials and process fragments currently defined for the system.

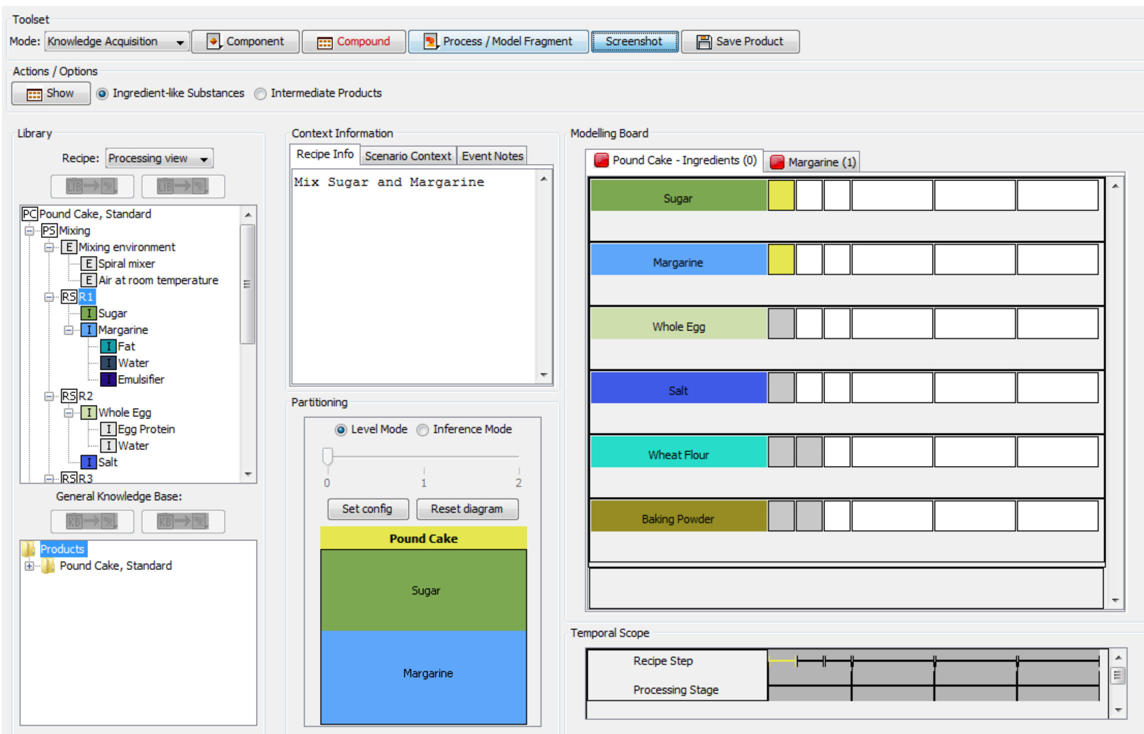

Figure 5. Workbench Overview with Compound Diagram and selected Process Context

Shown underneath is a repository for accessing knowledge objects from previously modeled systems for reuse in the current model; a main modeling board (right) which can display multiple, tabbed compound and scenario diagrams, as well as other tools for entering a new, decomposed recipe or defining details of newly specified materials. The small text-based tool in between the modeling board and the library is for retrieving or adding contextual information about the process system such as process contexts, recipe instructions, or notes to be stored for specific temporal events. Underneath this one is a tool for specifying the physical partitions of materials and for denoting when and how these partitions change during processes. The button tool bar at the top provides options for controlling the knowledge objects relevant for knowledge acquisition, and offers further commands specific to the different diagrammatic tools.

2.3.2 Knowledge Acquisition Procedure

Figure 5 shows the workbench in an initial state, after a decomposed recipe has been entered and two compound diagrams generated.[7] The figure also shows that the materials “Sugar” and “Margarine” have been selected as a process context (first recipe step) in the current diagram.

A user can, at any stage, extend the initial structural configuration by adding materials and relevant sub-components to the recipe, which will then appear in the library and the compound diagram. Visual heuristics facilitated by the compound diagram allow a quick perception about which materials are present in the system at identified time periods of processing, which are subject to the same activities, and therefore likely to interact (those within the same recipe step column – see highlighted temporal period in *Figure 5*). It also shows which materials are exposed to a similar environmental configuration (those within the same processing stage column) and hence, may be subject to any effects related to external materials, or energy applied.

A user would normally start knowledge acquisition for a given recipe by selecting related materials from the compound diagram matrix, and then creating a series of related process fragments (*Figure 3*). The fragments will be automatically associated to the temporal period within which the related contextual materials exist, without requiring an explicit start- or end time to be specified at this point. Each scenario diagram allows denoting the changes to the properties of materials or the system in the course of their interactions in the context of a single scenario (in order to minimize possible contradictions). Every property change made in the diagram also introduces an event mark on the discrete time line associated to each fragment (temporal scope) which can be annotated with quantitative values, if known. This can later be exploited in the inspection of process fragments which were specified within the same temporal context, by searching for quantitative values that allow a more precise alignment of these fragments on a joint temporal scale. Additional components can be added from the library or the knowledge objects repository to introduce additional materials, or describe related effects at different scales, such as corresponding changes at higher levels of structural abstraction.

Generally, selecting meaningful components within the single tools provided by the workbench activates any corresponding knowledge objects in other related tools (indicated by a similar color, for example), allowing the specification of different knowledge aspects about the same component or its related process. For instance, left-clicking a material in the scenario diagram also activates its representation in the partitioning tool which allows the specification of changes to the physical structure of the materials in the fragment. Also, changes to the structural configuration of the system are updated in the library view, extending the ontology of components and process fragments. Such changes are also reflected in the compound diagram such that materials are visually available (or no longer available from a certain temporal period onwards, respectively) for selection as contexts for the specification of further process fragments.

Since the interactive features of the workbench directly contribute to facilitating the knowledge acquisition method as a whole, this section can only indicate which individual tools, operations and visual heuristics are available overall. However, there are some video tutorials available which demonstrate the interactive features and their utility for knowledge acquisition as they are used in concert. These are accessible at the URL `http://tinyurl.com/q7b6fup` (link to a YouTube playlist).

[7] The top-level compound diagram is the one which is visible in *Figure 5*. A lower-level diagram created for the material “Margarine” is indicated but hidden behind the tab in focus.

2.4 Evaluation Summary

An initial evaluation was carried out by means of an industrial case study in the domain of food processing, using a semi-structured observational method.[8] The participating experts presented an ideal knowledge source for this study as they provided world-class expertise across different related scientific areas, including physical chemistry and biochemistry, and also in various industry-specific knowledge areas such as product development and design, process optimization, and knowledge transfer. The purpose of this evaluation was to find initial evidence indicating: 1) to what extent knowledge could be encoded and elicited with the tools and which cognitive operations supported, 2) the general usefulness of the single tools, combinations of their use, and range of their utilization, 3) the suitability of the generic model paradigm for accommodating knowledge about the different specific processes, 4) the adequacy of the representations for encoding knowledge, and 5) the suitability of the devised knowledge acquisition method in terms of the knowledge operations devised, and their sequencing.

The experts were asked to use the CMD SUITE workbench to specify knowledge about the processes occurring in the making of different food products. Two main sessions were held to address two kinds of products which were not previously examined during the development of the system, and which involve fundamentally different processes during their manufacture: A plain unit cake with high-ratio characteristics and a white bread based on a no-time dough.[9] Each session was just over four hours in duration, with breaks scheduled every 1-1,5h. Pen and paper were made available in case the participants wanted to use them.

Given some limited assistance, the experts were able to utilize the workbench successfully. In the course of the performed trials, 22 process fragments were specified and a substantial amount of knowledge was elicited. Most processes were of physical, chemical, or biochemical nature, and specified as the *solution*, *hydration*, *dispersion*, *coagulation*, *evaporation*, or *melting* of materials, for example. However, experts also used the diagrammatic tools to specify knowledge about more complex interactions, involving multiple types of processes, such as those related to *enzyme activity*, *yeast action*, *gluten formation*, or *the gelatinization of starch*. Some further notable insights from this study include:

- All knowledge expressed could be captured with the tools and representations provided by the software, without any need for additional, alternative means.
- The experts had only minor difficulties in using the workbench for performing knowledge acquisition tasks, given occasional guidance on how to encode particular types of knowledge.
- The ease with which the tools provided access to low-level process information enabled experts to spot gaps in their understanding of processes and begin considering higher-order relationships between processes rather than just the need to specify further any missing intra-process knowledge.

---

[8] The section gives only a very brief overview of the results. For a comprehensive account of the case study, we refer to Grau (2009), pp.157-182.

[9] *High ratio cakes* are made with a larger proportion of sugar in relation to the amount of flour used and can therefore carry more liquid than *low-ratio* cakes. *No-time dough* refers to the use of methods that speed up the proofing process in bread making, for instance by changing the properties of the dough chemically, or by increasing the amount of mechanical energy applied during mixing.

- Experts were able to consider the functionality and temporal sequence of partial processes at suitable levels of granularity, and able to relate effects to several levels of structural, behavioral, or functional levels of abstraction. They considered this very important for understanding the relationships between the processes in the system and the possible synthesis of new processes.
- Despite their prototypical state of design, the tools and their underlying representations clearly enabled different reasoning processes and the use of heuristics for knowledge acquisition and modeling. The representations provided a rich context for the experts to consider the workings of processes in new ways, on occasion (*Yeast action*) even to the extent that they perceived the utility of the conventional, mathematical approaches they normally used as inferior to the expressive capabilities of the diagrams.

In particular, the diagrams seemed to have triggered different reasoning processes, such as the reconceptualization of existing knowledge within a current process context. For example, the experts reconsidered the low-level interactions of the *Starch Gelatinization process* with other existing materials and processes and the effects on establishing the protein matrix of a product and hence, the higher-level effects on the structural development of the overall product (Grau, 2009, p.169). Further, two additional knowledge acquisition heuristics were discovered in connection with the scenario diagram, which was not anticipated. Here, the experts inspected previously specified process fragments either to 1) verify the presence of a certain material required for the current fragment, and 2) to check that the changes specified in the current process scenario were consistent with another, which they specified earlier (ibid., p.172).

Overall, the developed knowledge representations appeared appropriate and sufficiently flexible for encoding ill-structured process knowledge about the domain. Further, the individual operations outlined in the knowledge acquisition method seemed conceptually appropriate, arranged in the right order, and easily accessible to be performed iteratively by means of the different diagrammatic tools. We thus conclude that the properties of the expert knowledge did largely match the representational properties of the developed framework. Observations of the reasoning activities of the experts also showed the potential of diagrammatic representations to incorporate additional knowledge which had not been explicitly specified during their construction.

3. Discussion

In this section, we summarize the approach, considering its scope and limitations; indicate implications for related work; and outline future applications.

3.1 Scope and Limitations

Bodies of ill-structured process knowledge present a major problem for human cognition because performing the operations that underpin high-level cognitive tasks for knowledge acquisition, modeling, and discovery is an ill-structured problem. This problem has so far seen only little support by means of conceptual and computational developments because the inherent complexity and diverse dynamics that characterize domains with ill-structured process knowledge prevent its decomposition into largely independent sub-problems that can be examined and analyzed individually. We contend that the support of high-level cognitive tasks in such domains requires a holistic approach that must embrace their inherent complexity and reflect this in the integration and representation of the heterogeneous knowledge involved, and also in the problem spaces and methods that need to be developed for performing systematic knowledge operations.

The approach described in this paper has delivered a conceptual synthesis of representations and tools integrated within a framework for the decomposition and representation of generic model entities and their temporal and spatial context; the building of model structures in a compositional fashion; and the creation of knowledge fragments, based on process scenarios, which encode significant structural, behavioral, and functional configurations, their properties, and related changes. The CMD SUITE software workbench presents a prototype for demonstrating a human-driven, interactive approach to knowledge acquisition in domains with ill-structured process knowledge. The system provides the means for the flexible specification of knowledge at different degrees of granularity and resolution which allows the construction of knowledge-rich model structures considering many levels of abstraction, and multiple perspectives.

In particular, the approach builds on a characterization of different classes of ill-structured process knowledge (CMD processes) and has defined general notations that exploit the expressive power and epistemic benefits of diagrams to integrate the various concepts, notations, and models related to the multiple scientific paradigms underpinning the knowledge in the domain. Different types of novel representations have been developed to work in concert, in order to handle knowledge about the structural, behavioral, and functional aspects of materials and processes, and represent this knowledge on multiple temporal and spatial scales.

The approach has also outlined an overarching problem space for performing knowledge acquisition tasks, considering the context of knowledge modeling and discovery as related high-level tasks. The knowledge acquisition problem space comprises methods developed for recipe decomposition, heuristics for the identification of relevant process contexts, as well as the means for specifying and inspecting process fragments. The approach grants humans a high degree of control during the conduction of knowledge operations and allows these to be carried out in an iterative fashion, based on a system of structured, interactive representations and related methods and heuristics, which are embedded in an interactive software workbench.

The tools and the underlying framework were evaluated in a case study with industry experts, which provided many positive indicators for their utility in the target domain. It was found that the implemented methods were sufficient for the experts to provide detailed knowledge about manufacturing processes, identify complex relations between processes, and identify gaps in their current understanding based on the knowledge represented by the system. However, the development of more representations and methods is needed to extend the representational and computational capabilities of the system; for instance, to create better tools for the inspection and grouping of knowledge specified in individual process fragments to facilitate the exploration of higher-order facts and relations. For example, based on the current developments, higher-order representations can be developed that group process fragments when they 1) involve the same materials, 2) change similar properties, 3) share the same resources, 4) provide required inputs or outputs, respectively, 5) have components that share physical boundaries, 6) exist within the same temporal context. Further, the approach needs to be applied in additional domains and for different kinds of processes in order to further improve and evaluate its general applicability. While the designed diagrammatic representations may be applicable to various other domains, such as those which incrementally compound or decompose materials in the course of processes, their relevance may be limited in ill-structured process domains governed by slightly different conceptual dimensions. For instance, an integration of biological models based on regulatory signals will most likely require other representations to be developed.

Supporting high-level cognition for bodies of ill-structured process knowledge is a problem of tremendous scope and the capabilities of a single research program to develop comprehensive results are naturally limited. However, our characterization of the problem, the design of the representational system, the developed theoretical framework and the methods for knowledge acquisition, the software implementation, and the results of the first evaluation together present a first step towards the development of a more comprehensive approach to this important problem.

3.2 Implications for Related Research and Future Applications

The approach was informed by developments and results from many other scientific and technological areas, such as Cognitive Science and Artificial Intelligence, Computer Science, HCI, Information Visualization, Philosophy and History of Science, Knowledge Engineering, Process Modeling, and the Semantic Web.

For example, related work in Computational Scientific Discovery (e.g., Langley et al., 1987; Langley, 1998; Langley, 2000; Shrager, 2007) has been utilized and extended by the development of a problem space for the acquisition of ill-structured process knowledge. The approach facilitates the initial construction of process scenarios as partial states of a system of processes which may later be parameterized and explored with established methods of computational scientific discovery. In this respect, the approach also extends existing work in Knowledge Engineering because the developed methods do not require a complete specification of generic knowledge structures in order to be used at all. This is usually a prerequisite for the application of any established methods for knowledge acquisition in more well-structured domains (e.g., Schreiber et al., 1999; Noy, Ferguson, Musen, 2000). Further, the approach has demonstrated that the limitations imposed on domain experts by the use of logical or mathematical expressions can be reduced to prevent the occurrence of knowledge acquisition bottlenecks that have been observed with more rigid methods. The experts were able to use the diagrammatic representations to express and combine different kinds of knowledge more naturally, without the need to learn complex knowledge engineering syntax first. The approach also extends earlier work on diagrammatic knowledge acquisition (Cheng 1996, Cheng, Cupit & Shadbolt, 2001). The design of novel diagrammatic representations in this research program allowed for the 1) encoding of complex aspects of knowledge and 2) provision of knowledge-level interfaces through which human experts can express this knowledge. The approach has explored and adapted different concepts and ideas from qualitative and compositional modeling (e.g., Forbus, 1984; Falkenhainer & Forbus, 1991) in its design of a system of representations for the integration of heterogeneous process knowledge. The current range of work in Cognitive Science, HCI, and Information Visualization which studies the nature of diagrammatic representations typically compares existing notations, or designs new bespoke diagrams or graphical interfaces for well-structured knowledge domains and problems (e.g., Blackwell and Green, 2003; Ranson and Cheng, 2005; Vicente and Rasmussen, 1992). Our approach differs in that it presents an attempt to develop general notations for a whole domain class that embrace its inherent complexity, specifically compositional product manufacturing processes in the current workbench prototype, and CMD processes more generally under the overall approach.

Various scientific and industrial areas involve ill-structured process knowledge and may therefore benefit from a further development of this approach. In particular, this refers to domains that involve complex artifacts which do not originate from human design processes; collect large amounts of heterogeneous experimental data; or more generally, aim to incorporate existing knowledge and new insights into established theories and models of their domain. Applying the representations and methods of the approach to research on genome interaction, for instance, could support the capture of knowledge about protein interactions, the higher-level functions of protein clusters, and the development of heuristics for the identification of such interaction systems that may be of particular interest for closer study. Another application field may be drug discovery, which may benefit from the modeling of multi-scale effects of combined substances within human organisms. There are a range of applications in industry, as the framework provides a novel approach to production process modeling and product development and has a potential to yield entirely new classes of products for the latter. Industrial applications are likely to involve complex products where the interaction properties of product components and the dynamics of change during processing are currently not comprehensively understood, and where this has to be compensated for by using craft knowledge, experiential knowledge and the application of approximations instead of precise measures.

Encoding knowledge in *interactive diagrammatic knowledge-based systems* may provide new ways to represent, model and explore observational data and information that are difficult to interpret and relate to existing knowledge structures in their traditional notations. We think it likely that extending this approach will demonstrate the feasibility of such syntheses and challenge researchers to extend the scope of specific methods, and combine tools and methods into more powerful approaches in order to tackle interdisciplinary domains and the related problems that are currently difficult to address.

Acknowledgements

The research described in this paper was funded by the EPSRC ICASE Food Processing Faraday Partnership with the Campden & Chorleywood Food Research Association (CCFRA), Chipping Campden, GL55 6LD, UK. Additional funding was provided by the European Commission (FP7, Ref: 261826). Expert knowledge for development and evaluation has been provided by the CCFRA and BakeTran Ltd., UK.

References

Bechtel, W., & Richardson, R.C., 1993. *Discovering complexity: decomposition and localization as strategies in scientific research.* Princeton University Press.

Blackwell, A., & Green, T., 2003. Notational systems – the Cognitive Dimensions of Notations Framework. In: Carroll, J. (Ed.), *HCI Models, Theories, and Frameworks: Toward an Interdisciplinary Science*. Morgan Kaufmann.

Cheng, P.C.-H., 1996. Scientific discovery with law encoding diagrams. *Creativity Research Journal 9 (2/3)*, 145–162.

Cheng, P.C.-H., Barone, R., 2007. Representing complex problems: A representational epistemic approach. In Jonassen, D.H. (Ed.), *Learning to Solve Complex Scientific Problems.* Lawrence Erlbaum Associates, Mahmah, N.J., pp. 97–130.

Cheng, P.C.-H., Cupit, J., & Shadbolt, N.R., 2001. Supporting diagrammatic knowledge acquisition: An ontological analysis of Cartesian graphs. *International Journal of Human Computer Studies, 54*, 457–494.

Cheng, P.C.-H., & Simon, H.A., 1995. Scientific discovery and creative reasoning with diagrams. In Smith, S., Ward, T., Finke, R. (Eds.), *The Creative Cognition Approach*. MIT Press, Cambridge, MA, pp. 205–228.

Falkenhainer, B., & Forbus, K.D., 1991. Compositional Modeling: Finding the Right Model for the Job. *Artificial Intelligence, 51*, 95–143.

Forbus, K.D., 1984. Qualitative Process Theory. *Artificial Intelligence, 24*, 85–168.

Grau, R.R., 2009. *The Acquisition and Representation of Knowledge about Complex Multi-Dynamic Processes*, Doctoral dissertation, Department of Informatics, University of Sussex, England. (http://www.academia.edu/attachments/31869487/download_file)

Jonassen, D.H., 1997. Instructional design models for well-structured and ill-structured problem-solving learning outcomes. *Educational Technology: Research and Development, 45 (1)*, 65–94.

Klahr, D., & Dunbar, K., 1988. Dual space search during scientific reasoning. *Cognitive Science, 12*, 1–48.

Langley, P., Simon, H.A., Bradshaw, G.L., & Zytkow, J.M., 1987. *Scientific Discovery: Computational Explorations of the Creative Process*. MIT Press, Cambridge, Massachusetts.

Langley, P., 1998. The computer-aided discovery of scientific knowledge. In *Proceedings of the First International Conference on Discovery Science*. Springer Verlag, Fukuoka, Japan.

Langley, P., 2000. The computational support of scientific discovery. *International Journal of Human-Computer Studies, 53*, 393–410.

Newell, A., & Simon, H.A., 1972. Human Problem solving. Prentice Hall, Englewood Cliffs, NJ.

Noy, N.F., Ferguson, R.W., & Musen, M.A., 2000. The knowledge model of Protégé-2000: combining interoperability and flexibility. In *Proceedings Second International Conference on Knowledge Engineering and Knowledge Management*, France.

Ranson, D., & Cheng, P.C.-H., 2005. Graphical tools for heuristic visualization. In: Kendall, G., Lei, L., & Pinedo, M. (Eds.), *Proceedings of the 2nd Multidisciplinary International Conference on Scheduling: Theory and Applications*, (pp. 658-667). New York, USA.

Schreiber, G., Akkermans, H., Anjewierden, A., De Hoog, R., Shadbolt, N., Van de Velde, W., Wielinga, B., 1999. *Knowledge Engineering and Management — The CommonKADS Methodology*, 2nd printing 2001. The MIT Press, Cambridge, Massachusetts, London, England.

Simon, H.A., 1962. The architecture of complexity. In *Proceedings of the American Philosophical Society*, No. 106. pp. 467–482.

Simon, H.A., 1973. The structure of ill-structured problems. *Artificial Intelligence 4 (3)*, 181–201.

Shrager, J., 2007. The evolution of BioBike: Community adaptation of a biocomputing platform. *Studies in History and Philosophy of Science, 38 (4)*, 642–656.

Vicente, K.J., & Rasmussen, J., 1992. Ecological Interface Design: Theoretical foundations. *IEEE Transactions on Systems, Man and Cybernetics, 22*, 589–606.