Fondazione Bruno Kessler
Institute of Research on Science and Technology
FBK-Irst

NICTA
Canberra Research Laboratory
and
The Australian National University

National Research Council of Italy
Institute for Cognitive Science and Technology
Planning and Scheduling Team
ISTC-CNR [PST]

Third International Competition on Knowledge Engineering for Planning and Scheduling

ICKEPS 09

Thessaloniki, Greece, September 20, 2009
(In conjunction with ICAPS 09)

Report of the Board of Judges

October 2009

Piergiorgio Bertoli
FBK
bertoli@fbk.eu

Adi Botea
NICTA and the ANU
adi.botea@nicta.com.au

Simone Fratini
ISTC-CNR
simone.fratini@istc.cnr.it

Contents

1 Introduction and scope of the document

The International Competition on Knowledge Engineering for Planning and Scheduling has been running since 2005 as a bi-annual event promoting the development and the use of knowledge engineering methods and techniques within planning and scheduling (P&S). Past events include ICKEPS-05 (held during ICAPS at San Francisco in June 2005), and ICKEPS-07 (held during ICAPS at Providence in September 2007).

While those events aimed at KEPS in general, in ICKEPS-09 special emphasis has been put on the evaluation of *"tools, translators and techniques that, when input with a model described in an application-area-specific language, output solver-ready domain models"* (see `http://kti.mff.cuni.cz/~bartak/ICKEPS2009/`).

The participants of ICKEPS-09 contributed with tools and translators from a wide range of application areas, such as:

- Web Services and Semantic Web
- Workflows and Business Process Modeling
- Databases and Data Mining
- UML
- E-Learning
- Human-Instructable Computing

This document constitutes the report of the board of judges that evaluated the proposals, choosing two winners after a careful evaluation of both the paper presentations and the demonstration of the tools. The aim of this report it is not only to explain the reasons behind the choice of the two winners, but also to provide feedback to all the competitors from three colleagues that took the time to analyze and test their tools.

2 Competition setting and evaluation criteria

ICKEPS-09 ran in two stages, in conjunction with the ICAPS-09 conference. First, in the pre-conference stage, the competitors submitted papers describing the tools, focusing in particular on the translation processes from their application domains to the chosen planning language (and possibly back). In this phase, a program committee of 15 members reviewed the papers with the main goals of assessing their appropriateness to the competition, evaluating the contributions and providing feedback to their authors.

Then, during the conference, the competitors gave talks about their systems in a workshop-like setting that lasted a morning. Finally, in the afternoon of the same day, they presented the systems during a plenary demo session, open to the public. During this session the judges evaluated and tried the tools, interacting directly with the competitors that were running the demo.

During both phases, the judges continuously evaluated the contributions trying to assess the value of the proposed tools under two main general criteria, which reflect their potential to link planning and scheduling research to applicative areas in two directions:

- What advantages does the use of planning proposed in the tool bring in solving problems in the chosen application domain?

- What is the added value for the planning community in having the chance of using P&S technologies in the chosen application domain?

The demo session phase was particularly relevant to assess these criteria, as well as a number of slightly more specific criteria related to Software Engineering (SE) issues and P&S issues, which were published on the ICKEPS-09 page and we recap below:

- Portability, meant as a measurement of the "difficulty of using the tool out of the laptop of the competitor".
- Robustness, meant as a measurement of how much the quality of the translation is input-dependent.
- Usability, meant as a measurement of how much the tool is usable either by AI experts or target domain experts.
- Spread of use of the translator. (How many people have used the tool so far? To do what? How widely spread is the input language?)
- Perceived added value to the application area. This is meant to be an evaluation of the impact of the tool.
- Flexibility. How easy would it be to use the tool for domains out of those foreseen by the authors? How demanding is it to extend the set of problems that it is possible to translate?

More specific P&S criteria evaluated were:

- Originality. Is the approach original (with respect to previous proposals by other authors)?
- Comprehensiveness of the translation process. Is the sub-set of the input language/problems that can be translated well defined? Is the chosen subset broad enough for targeting significant problems in the application community?
- The challenges involved in the translation process. The more different the input/output languages are, the more valuable should the tool be.
- Connection with the P&S technology. Is the tool a comprehensive KE tool? Is there any available useful translation of an output plan or schedule back into the application domain? Is there any feature available to interact with the planner/scheduler?
- Ingenuity of the translator.
- The availability of solvers to handle the translated domain model and the performance of the planner and/or scheduler (when available) with the translated domains.

In particular, in the view of our main high-level criteria, the *spread of use* and the *challenges involved in the translation process* have been considered as driving factors for the evaluation purposes (while of course taking into account also the remaining criteria).

The judges proceeded to evaluate the above criteria in a way as accurate as possible, within the time frame of ICKEPS-09. It has to be remarked that, as such, the evaluation stands on the judges' subjective views, given the practical infeasibility of objectively measuring each criteria. In the following, this document recaps the results of the evaluation, listing the evaluated systems in the same order in which they appear in the ICKEPS-09 proceedings.

3 Evaluation of the competitors

3.1 LOCM: A Tool for Acquiring Planning Domain Models from Action Traces

Author: Stephen Cresswell

Description: LOCM is a system which carries out the automated induction of a planning domain, given sets of example action sequences described in a simple input language.

Application Domain: This is a general purpose tool, and may cover any task for which example sequences can easily be collected, e.g. by logging workflows or moves in a computer game.

Application Area Languages: So far does not consider application languages, but assumes a simple "action sequence language"

Planning Language: PDDL

Strengths:

- Has a wide (even though, so far, only *potential*) range of application areas.
- The underlying domain discovery mechanism is quite ingenious and not trivial.

Weaknesses:

- While many *potential* areas are suggested, there is no strong witness for applications so far.
- At the current stage, it requires strong specific expertise (of PDDL and possibly of LOCM) to actually make use of the extracted domain to specify problems in it.
- The results of the translation mechanism stand on some assumptions and depend on the available quantity of sample sequences. Whether the assumptions hold, and whether the results are accurate given the available samples, need be checked in practice on applicative domains.
- In general, the environment is not mature, at this stage, for immediate adoption outside the area of planning experts, and requires some insight on the tool itself.

Comments: LOCM has definitely one of the most interesting underlying ideas among the competitors. The process of information acquisition is not trivial and the approach suggests many interesting research directions. The biggest drawbacks are in the tool lack of maturity and in the lack of clearly defined application areas: LOCM is still at the stage of a proof of concept research prototype, more than being an engineer-ready tool, and there is not a specific target area of application to evaluate the added value of the approach.

3.2 On Compiling Data Mining Tasks to PDDL

Authors: Susana Fernández, Fernando Fernández, Alexis Sánchez, Tomás de la Rosa, Javier Ortiz, Daniel Borrajo, David Manzano

Description: An architecture and tool based on automated planning that helps users (not necessarily experts in data-mining) to perform data-mining tasks. In this architecture the data-mining processes are recasted as automated planning, which allows to automatize the data-mining knowledge flow construction.

Application Domain: Data Mining
Application Area Languages: Predictive Model Markup Language (PMML), Knowledge Flow for Machine Learning (KFML)
Planning Language: PDDL

Strengths:

- Clearly defined and highly relevant applicative area.
- Considerable perceived added value to the application area.
- Planned to be tested (and hopefully adopted) within a production setting.
- The translation process is comprehensive, and the sub-set of the input language/problems that can be translated is well defined.

Weaknesses:

- The results of this tool depend crucially on some "metric estimates" that are encoded within the representation of the data mining operations. At the current stage, such values are hard-coded and do not reflect any experience that the system might "learn" in time, when data mining is actually performed. In order for the tool to be fully fit for production, the addition of a feedback mechanism would be very relevant.
- The translation problem is not very challenging from a research perspective, as most of the problem stands in the (fixed, and not automated) encoding of data mining operations.
- In terms of flexibility, since the tool stands on a fixed encoding of data mining operations, it might require work and substantial expertise to adapt the system to work on an extended set of operations (even though the need for this is quite small, given the comprehensive coverage of the operations featured already).

Comments: This tool targets an important, well defined, and spread application area: data bases and data mining. The perceived added value of this tool in the application area is clear: to support engineers in defining workflows of data mining tasks according to some criteria, by taking advantage of planning technology to automate the construction of an optimal plan of data mining tasks. The two main drawbacks of this tool are (a) the lack of a feedback mechanism that would keep aligning the tool with the current performances of data mining actions, and (b) the fact that the tool stands on top of a fixed PDDL encoding of data mining actions. Especially the first, in our view, might put at harm the full integration of the tool in a production setting, since it requires continuous monitoring of results and manual realigning of estimates within the action encoding.

3.3 Modeling E-Learning Activities in Automated Planning

Authors: Antonio Garrido, Eva Onaindia, Lluvia Morales, Luis Castillo, Susana Fernández, Daniel Borrajo

Description: This tool compiles a course defined in a standard e-learning language into a planning domain, and a file containing student learning information into a planning problem. The learning designs are automatically generated from the produced plans. The application domain is translated into a planning domain using three approaches with an increasing level of complexity: hierarchical, PDDL-based using actions with conditional effects, and PDDL-based using durative actions.

Application Domain: E-learning

Application Area Languages: IMS-MD, IMS-LIP

Planning Language: HTN, PDDL

Strengths:

- Clearly defined applicative area.
- Perceived added value to the application area.
- The tool is comprehensive, and considers several planning techniques to cover different varieties of the problem presented in the application domain.
- The tool, featuring a well-developed interface, appears mature enough for usage by "standard users" which are not expert (nor aware) of planning languages and techniques.

Weaknesses:

- The applicative area is quite specific, and the impact of the tool on the area should be tested more in depth by having e.g. a set of universities or departments adopt the tool, and provide feedback on it (based, in turn, on the feedback of students as well as other actors).
- The translation mechanisms, as they are presented, do not appear extremely challenging from a research perspective as, in most cases, they lead to planning domains quite similar to (excerpts of) the inputs.

Comments: This tool provides a good functionality for supporting users in the target area. The area appears to be quite specific, and in our view the tool would require some more substantial testing to make sure that the proposed techniques are adopted in full by the audience in the area. Tests in this sense might reveal e.g. the need for some feedback mechanism to cope with cases where users may want to customize the proposed list of courses to some extent.

3.4 JABBAH: A Java Application Framework for the Translation Between Business Process Models and HTN

Authors: Arturo González-Ferrer, Juan Fernández-Olivares, Luis Castillo

Description: JABBAH is a tool supporting the analysis of resource allocation over BPMN-specified workflows by means of planning techniques. It provides an extensible framework where a workflow graph is first transformed into an equivalent nested process model, which is subsequently mapped into a P&S domain and problem expressed in HTN-PDDL. The solution to such a problem provides a suitable allocation of resources for the given input workflow.

Application Domain: Business Process Modeling

Application Area Languages: Businness Process Management Notation, XML Process Definition Language (BPMN/XPDL)

Planning Language: HTN

Strengths:

- Clearly defined and highly relevant applicative area.
- Considerable perceived added value to the application area.
- The tool, which includes a user-friendly interface, appears mature for usage by "standard users" which are not expert (nor aware) of planning languages and techniques, and fully supports them in all the phases of the analysis, from the input using BPMN/XPDL standards to a clearly understandable output schedule.
- The underlying translation of the problem as a P&S problem is not trivial and is interesting research-wise.

Weaknesses:

- The tool may be improved for what concerns flexibility: at the current stage, for instance, it does not allow easily specifying complex criteria for preferable (or vice versa, undesirable) allocations.
- Its strengths would be better assessed by a more thorough testing on large-scale workflows (even though current test scenarios are not trivial either).

Comments: The JABBAH system provides a neat tool for analysts that need to perform resource allocation analysis on business workflows, embedding a non-trivial transformation of BPMN-expressed workflows in terms of HTNs. By providing a fully automated support of the analysis, allowing engineers to exploit the vastly diffused BPMN standard for workflow specification, and neatly presenting the results, this system may appeal a very wide and relevant audience. Henceforth, JABBAH may have a considerable potential impact outside the planning community. Furthermore, the system appeared solid and well usable, and in general scored well under the criteria considered in ICKEPS'09. *For these reasons, JABBAH was chosen as one of the winners of ICKEPS-09.*

3.5 PORSCE II: Using Planning for Semantic Web Service Composition

Authors: Ourania Hatzi, Georgios Meditskos, Dimitris Vrakas, Nick Bassiliades, Dimosthenis Anagnostopoulos, Ioannis Vlahavas

Description: PORSCE II is an integrated system that performs automatic semantic web service composition through planning. The web service composition problem is translated into a planning problem in PDDL. In the translation process, semantic "proximity" information is considered according to user-provided preferences, so that solutions of different degree of flexibility can be produced. The planning problem is then solved using external domain-independent planning systems, and the solutions are visualized and ranked according to their quality.

Application Domain: Web Service Composition

Application Area Language: Semantic OWL (OWL-S)

Planning Language: PDDL

Strengths:

- Tackles an interesting applicative area which is evolving fast, providing a potentially crucial functionality.
- Well engineered, it provides effective support to standard users (non-expert and unaware of planning techniques) for solving service composition tasks.

Weaknesses:

- Lacks support for the actual deployment of discovered service compositions.
- While allowing for flexible compositions with approximate concept matching is an interesting idea, the extend to which such compositions can be of use has to be checked in practice, on a real work-bench of use cases.
- Based on semantic similarity, the system is able to propose one-to-one component substitutions. This feature has practical appeal (e.g., to solve "unavailable service" instances), but it ought to be expanded to cover general one-to-many substitutions.

Comments: This tool is quite interesting and well designed, and tackles a relevant problem in a fast-growing area, implementing an interesting idea of "performing composition based on approximate concept matching". At the current stage, PORSCE II suffers from two main drawbacks. First, it fails to provide full-cycle support, leaving users with the burden of self-developing and deploying an (automatically discovered) composite service. This functionality is probably not difficult to design and embed in the system, but it is very relevant for the usability of the tool within a production setting. Second, the tool does not fully exploit P&S technology to give the possibility of substituting failing services with non trivial combinations of other available services, being limited to simple one-to-one substitutions. Finally, a better assessment of the practical impact of the "approximate matching" idea is recommendable.

3.6 Augmenting Instructable Computing with Planning Technology

Authors: Clayton T. Morrison, Daniel Bryce, Ian R. Fasel, Antons Rebguns

Description: This complex tool proposes the enhancement of an environment for instructable computing by means of planning techniques that, integrated with some learning patterns, allow to automatically discover concepts and procedures in the face of incompletely described teaching lessons.

Application Domain: Human-Instructable Computing

Application Area Languages: MABLE's Interlingua (IL)

Planning Language: PDDL

Strengths:

- Conceptually very interesting and novel idea, integrating several different aspects (monitoring of activities, learning patterns, encoding of teaching as planning) to significantly enhance an already advanced environment.
- Potentially useful to several specific areas (all those areas where instructable computing mechanisms may be adopted).

Weaknesses:

- At the current stage, the tool requires significant and very specific know-how for its usage, and has not been linked to specific application interfaces.
- At the current stage, the tool is not fully integrated with P&S techniques, whose application takes place on the basis of manual intervention.
- The potential ability of the tool to enhance instructable computing ought to be assessed on the field by testing it on a wealth of use cases.

Comments: The complexity of this tool, and the fact that it proposes a technical interface to it, made its evaluation significantly difficult. The general idea of the tool, and the fact that it integrates a variety of different concepts to achieve the automated enhancement of an instructable computing environment, are clear winning points of this proposal. However, at the current stage of development, the tool is more suitable for experts that have a strong know-how on the tool's (complex) input language and workings. Furthermore, the tool is not yet completely integrated with the planning technology it shall exploit.

3.7 From Requirements and Analysis to PDDL in itSIMPLE3.0

Authors: Tiago Stegun Vaquero, José Reinaldo Silva, Marcelo Ferreira, Flavio Tonidandel, J. Christopher Beck

Description: it-SIMPLE is a long-standing project dedicated to support engineers and researchers alike in the design phases of applications in terms of planning domains, to obtain a sound input-ready model for planners. With this tool it is possible to translate requirements in UML into a solver-ready PDDL model, to exploit a variety of planning technologies, and to analyze the results.

Application Domain: Not specified

Application Area Languages: Unified Modeling Language (UML)

Planning Language: PDDL

Strengths:

- A comprehensive tool supporting users in full, from the design of domains via UML, to the exploitation of planners on them, to the analysis of results.
- The tool appears remarkably robust and mature, featuring easy-to-grasp interfaces.
- Provides added value both for the P&S community, and for engineers intending to exploit P&S techniques in their areas.
- Even though not oriented to a specific applicative area, has been successfully used in applicative contexts.

Weaknesses:

- Does not fit perfectly with the focus of the competition (not "application-specific").
- While it provides a big weaponry of functions, some more are desirable (e.g., a way to establish a hierarchy of UML models to deal with very large-scale projects), and some existing ones need improvement (most substantially, Petri net modeling and analysis).

Comments: The itSIMPLE 3.0 system showed as a prominently robust and comprehensive system capable to effectively support engineers and scientists in modeling domains, planning on them and analyzing the outcomes of planning activities, by means of user-friendly GUI interfaces and taking the well-known UML standard as the key representational means. The system encompasses a variety of functionality and interfaces, and clearly incorporates a large wealth of work carried out since its inception, several years ago. While not focused on a specific applicative area, the tool has been exploited in several applicative fields, witnessing the strength of this workbench and its potential to significantly widen the forum of the users of planning techniques. *For these reasons, itSIMPLE3.0 was chosen as one of the winners of ICKEPS-09.*