Planning in the Wild: Modeling Tools for PDDL*

VOLKER STROBEL† ALEXANDRA KIRSCH‡

Writing and maintaining planning problems, specified in the widely used *Planning Domain Definition Language* (PDDL) can be difficult, time-consuming, and error-prone. One reason seems to be the missing support by engineering tools. The present study proposes MYPDDL – a modular toolkit for developing and manipulating PDDL domains and problems. To evaluate MYPDDL, we compare it to existing knowledge engineering tools for PDDL and experimentally assess its usefulness for novice PDDL users.

1 INTRODUCTION

A large community of researchers dedicate their efforts to Artificial Intelligence (AI) planning. However, the process made in this community is often ignored when it comes to developing complete AI systems. Planning is a fundamental cognitive function that is useful for most systems claiming to be intelligent, such as autonomous robots or decision support systems. This raises the question why planning is not used in more systems. We believe that one reason is the gap between modeling textbook toy problems and modeling complex, real-world problems.

The standard AI planning language PDDL (*Planning Domain Definition Language*) differentiates between domain files with definitions of types, predicates and actions, and problem files with definitions of objects and goals. Realistic scenarios contain hundreds of objects, different agents with different capabilities, able to perform a large variety of actions. Modeling such worlds soon gets confusing.

This problem is not specific to planning, but poses a challenge to software engineering in general. As projects grow in size, developers have to be supported with appropriate tools to keep track of the overall structure.

This paper proposes MYPDDL, a set of tools for modeling large PDDL domains and associated problems. Section 2 discusses existing tools for PDDL. Section 3 presents MYPDDL's modules and design principles. They are evaluated with a user test in Section 4. Section 5 concludes with an outlook on further steps necessary to improve the availability of planning for intelligent system development.

2 RELATED WORK

There have been some attempts to provide modeling tools for PDDL. This section introduces the three most sophisticated tools we found.

PDDL STUDIO [1] is an application for creating and managing PDDL projects, i.e. a collection of PDDL files. PDDL STUDIO's integrated development environment (IDE) was inspired by Microsoft Visual Studio and imperative programming paradigms. Its main features are syntax highlighting, error detection, context sensitive code completion, code folding, project management, and planner integration. PDDL STUDIO's error detection can recognize both syntactic (missing keywords, parentheses, etc.) and semantic (wrong type of predicate parameters, misspelled predicates, etc.) errors.

*The final publication is available at Springer via http://dx.doi.org/10.1007/978-3-319-11206-0_27

†Contact email address: volker.strobel87@gmail.com

‡University of Tübingen

A major drawback of PDDL STUDIO is that it is not updated regularly and only supports PDDL 1.2. Later PDDL versions contain several additional features such as durative actions, numeric fluents, and plan metrics [2].

ITSIMPLE [3] follows a graphical approach using Unified Modeling Language (UML) diagrams. In the process leading up to ITSIMPLE, UML.P (UML in a Planning Approach) was proposed, a UML variant specifically designed for modeling planning domains and problems [4].

ITSIMPLE's modeling workflow is unidirectional as changes in the PDDL domain do not affect the UML model and UML models have to be modeled manually, meaning that they cannot by generated from PDDL. However, Tonidandel, Vaquero, and Silva [5] present a translation process from a PDDL domain specification to an object-oriented UML.P model as a possible integration for ITSIMPLE. This translation process makes extensive semantic assumptions for PDDL descriptions. For example, the first parameter in the `:parameters` section of an action is automatically declared as a subclass of the default class `Agent`, and the method is limited to predicates with a maximum arity of two. The currently version of ITSIMPLE does not include the translation process from PDDL to UML.

Starting in version 4.0, ITSIMPLE expanded its features to allow the creation of PDDL projects from scratch (i.e. without the UML to PDDL translation process) [6]. Thus far, the PDDL editing features are basic. A minimal syntax highlighting feature recognizes PDDL keywords, variables, and comments. ITSIMPLE also provides templates for PDDL constructs, such as requirement specifications, predicates, actions, initial state, and goal definitions.

Both PDDL STUDIO and ITSIMPLE do not build on existing editors and therefore cannot fall back on refined implementations of features that have been modified and improved many times throughout their existence.

PDDL-mode[1] for the widely used Emacs editor builds on the sophisticated features of Emacs and uses its extensibility and customizability. It provides syntax highlighting by way of basic pattern matching of keywords, variables, and comments. Additional features are automatic indentation and code completion as well as bracket matching. Code snippets for the creation of domains, problems, and actions are also available. Finally, PDDL-mode keeps track of action and problem declarations by adding them to a menu and thus intending to allow for easy and fast code navigation.

PDDL-mode for Emacs supports PDDL versions up to 2.2, which includes derived predicates and timed initial predicates [2], but does not recognize later features like object-fluents.

In sum, there is currently no tool available supporting all features of PDDL 3.1, nor all the steps in the modeling process.

3 MYPDDL

MYPDDL is designed as a modular framework. We first introduce the implemented modules and then explain their details with respect to design guidelines for knowledge engineering tools.

3.1 *Modules*

MYPDDL-IDE is an integrated development environment for the use of MYPDDL in the text and code editor *Sublime Text*[2]. Since MYPDDL-SNIPPET and -SYNTAX are devised explicitly for Sublime Text, their integration is implicit. The other tools can be used independently of Sublime Text with the command-line interface and any PDDL file, but were also integrated into the editor.

[1] http://rakaposhi.eas.asu.edu/planning- list-mailarchive/msg000085.html
[2] `http://www.sublimetext.com`

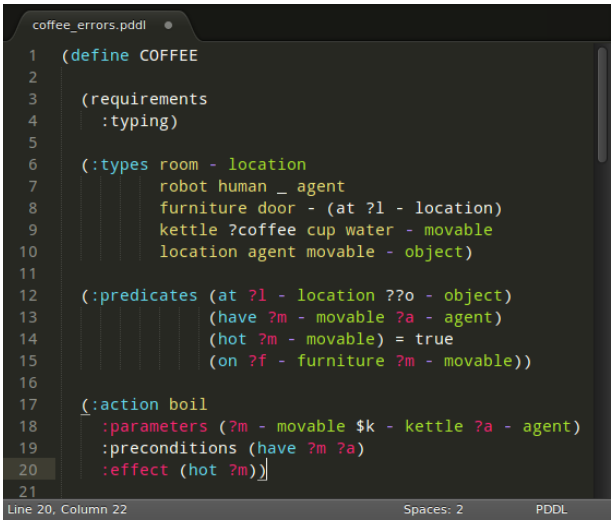

Figure 1: Syntax highlighting using myPddl-ide. White text contains errors.

mypddl-syntax is a context-aware syntax highlighting feature for Sublime Text. It distinguishes all pddl constructs up to version 3.1. Using regular expressions that can recognize both the start and the end of code blocks by means of a sophisticated pattern matching heuristic, myPddl-syntax identifies pddl code blocks and constructs and divides them into so called scopes, i.e. named regions. Sublime Text colorizes the code elements via the assigned scope names and in accordance with the current color scheme. These scopes allow for a fragmentation of the pddl files, so that constructs are only highlighted if they appear in the correct context. Thus missing brackets, misplaced expressions and misspelled keywords are visually distinct and can be identified (see Figure 1).

mypddl-new helps to organize pddl projects by generating the following folder structure:

```
project-name/
├── domains/
├── problems/
│   └── p01.pddl
├── solutions/
├── domain.pddl
└── README.md
```

The domain file `domain.pddl` and the problem file `p01.pddl` initially contain corresponding pddl skeletons which can also be customized. All problem files that are associated with one domain file are collected in the folder `problems/`. `README.md` is a Markdown file, which is intended for (but not limited to) information about the author(s) of the project, contact information, informal domain and problem specifications, and licensing information. Markdown files can be converted to html by various hosting services (like GitHub or Bitbucket).

mypddl-snippet provides code skeletons, i.e. templates for often used pddl constructs such as domains, problems, type and function declarations, and actions. They can be inserted by typing a triggering keyword.

mypddl-clojure provides a preprocessor for pddl files to bypass pddl's limited mathematical capabilities, thus reducing modeling time without overcharging planning algorithms. We decided to use Clojure [7], a modern Lisp dialect that runs on the Java Virtual Machine (jvm) [8], facilitating input and output of the Lisp-style pddl constructs.

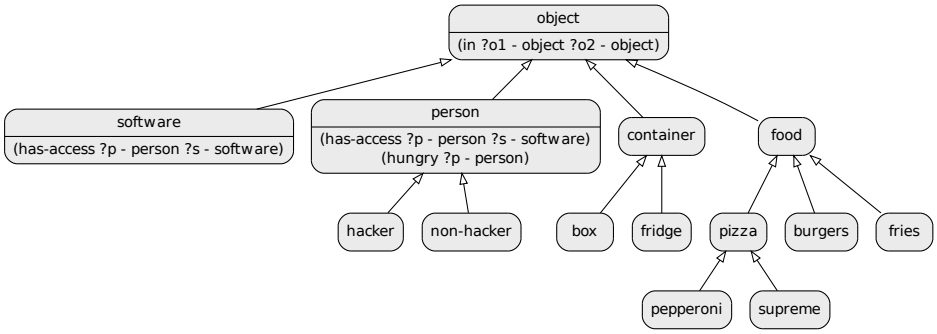

Figure 2: Type diagram generated by myPddl-diagram

mypddl-distance provides special preprocessing functions for distance calculations. For domains with spatial components, the distance of objects is often important and should not be omitted in the domain model. However, calculating distances from coordinates requires the square root function, which is not supported by pddl (it only supports the four basic arithmetic operators). More sophisticated calculations can be achieved with the supported operators, but the solutions are rather inefficent and inelegant [9]. By calculating the distances offline and including them as additional predicates in the problem file using myPddl-distance, the distances between objects are given to the planner as part of the problem description.

mypddl-diagram generates a png image from a pddl domain file as shown in Figure 2. The diagrammatic representation of textual information helps to quickly understand the connection of hierarchically structured items and should thus be able to simplify the communication and collaboration between developers. In the process of generating the diagrams, a copy of the pddl file is created, so that a simple version control is also included.

3.2 *Design Principles*

As guidelines for design decisions, we used the seven criteria for knowledge engineering tools proposed by Shah et al. [10] as well as general usability principles. Operationality instantiates, whether the generated models can improve the planning performance. This is not a design principle for myPddl, because we assume that myPddl does not improve the quality (with respect to planning performance) of the resulting pddl specifications. Therefore, we replaced this criterion with functional suitability from the iso/iec 25010 standard, which is defined as the degree to which the software product provides an appropriate set of functions for specified tasks and user objectives (iso 25010 6.1.1). myPddl supports the current version 3.1 of pddl. It encompasses and exceeds most of the functionality of the existing tools. It specifically supports basic editor features with a high customizability as well as visualization support. Collaboration: With the growing importance of team work and team members not necessarily working in the same building, or in the same country, there is an increasing need for tools supporting the collaboration effort. In developing myPddl, this need was sought to be met by myPddl-diagram. Complex type hierarchies can be hard to overlook, especially if they were constructed by someone else. Therefore, a good way of tackling this problem seemed to be by providing a means to visualize such hierarchies in the form of type diagrams.

experience : myPddl was designed specifically for users with a background in AI, but not necessarily in pddl. The tools are similar to standard software

engineering tools and should thus be easily learnable. The user evaluation (Section 4.2) confirms that myPddl helps novices in pddl to master planning task modeling. In addition, it is also possible to customize myPddl so as to adapt its look and feel to other programs one is already familiar with, or simply to make it more enjoyable to use. The project site[3] provides myPddl video introductions and a manual to get started quickly.

efficiency : All myPddl tools are intended to increase the efficiency with which pddl files are created. myPddl-snippet enables the fast creation of large and correct code skeletons that only need to be complemented. myPddl-syntax can reduce the time spent on searching errors. Code folding allows users to hide currently irrelevant parts of the code and automatic indentation increases its readability. To easily keep track of all the parts of a project, folders are automatically created and named with myPddl-new. myPddl-clojure and -distance allow for a straightforward inclusion of numerical values in the problem definition.

debugging : myPddl -syntax highlights all syntactically correct constructs and leaves all syntactical errors non-highlighted. In contrast, pddl-mode for Emacs and itSimple only provide basic syntax highlighting for emphasizing the structure. pddl studio explicitly detects errors, but the user is immediately prompted when an error is detected. Often, such error messages are premature, for example, just because the closing parenthesis was not typed yet, does not mean it was forgotten. myPddl indicates errors in a more subtle way: syntactic errors are simply not highlighted, while all correct pddl code is. The colors are customizable, so that users can choose how prominent the highlighting sticks out.

maintenance : The possibility to maintain pddl files is a key aspect of myPddl. The automatically generated type diagram (myPddl-diagram) gives an overview of the domain structure and thereby serves as a continuous means of documentation. Helping to understand foreign code, though, it follows logically that myPddl-diagram also helps in coming back and changing ones own models if some time has elapsed since they were last edited. The basic revision control feature of myPddl-diagram keeps track of changes, making it easy to revert to a previous domain version. Furthermore, myPddl-new encourages adhering to an organized project structure and stores corresponding files at the same location. The automatically created readme file can induce the user to provide further information and documentation about the pddl project. Support: myPddl-ide can be installed using Sublime Text's Package Control[4]. This allows for an easy installation and staying up-to-date with future versions. In order to provide global access and with it the possibility for developing an active community, the project source code is hosted on GitHub[5]. Additionally, the project site provides room for discussing features and reporting bugs.

4 validation and evaluation

To assess the utility of myPddl, we used the criteria listed in Section 3.2. The functional suitability was evaluated using a benchmark validation, comparing myPddl s functionality with the tools described in Section 2. The criteria collaboration, experience, efficiency, and debugging were evaluated in a user test. The myPddl components supporting maintenance are the same ones that are used in the user test, but their long-term usage is difficult to evaluate. The support criterion depends primarily on the infrastructure, which has been established as explained in 3.2.

[3] `http://pold87.github.io/myPDDL`
[4] `https://sublime.wbond.net/about`
[5] https://github.com/Pold87/myPDDL

4.1 *Benchmark Validation*

Functional suitability encompasses the set of functions to meet the user objectives. The tools of Section 2 basically all follow the same objectives as myPddl: creating pddl domains and problems. The features offered by each tool are summarized in Table 1. Besides supporting the latest pddl version, a strength of myPddl is its high customizability, which comes with the Sublime Text editor. Being the only one of the four tools capable of visualizing parts of the pddl code, it must be understood as complementary to itSimple, which takes the opposite approach of transforming uml diagrams into pddl files. The fact that myPddl does not check for semantic errors is not actually a drawback as planners will usually detect semantic errors. All in all, myPddl combines the most useful tools of pddl studio, itSimple, and pddl-mode for Emacs and strives to support the planning task engineer during all phases of the modeling process. Additionally, it features some unique tools, such as domain visualization and an interface with a programming language. It can therefore be concluded that myPddl provides an appropriate set of functions for developing pddl files and is thus functionally suitable.

4.2 *User Evaluation*

The two most central modules of myPddl are myPddl-syntax and myPddl-diagram, since they support collaboration, efficiency, and debugging independently of the users experience with pddl. To evaluate their usability, they will be evaluated in a user study.

4.2.1 *Procedure*

We invited eight participants[6] to a user test (three female, average age 22.9, standard deviation 0.6), who had some basic experience with at least one Lisp dialect (in order not to be confused with the many parentheses), but no experience with pddl or AI planning in general.

No earlier than 24 hours before the experiment was to take place, participants received the web link to a 30-minute interactive video tutorial on AI planning and pddl. This method was chosen in order not to pressure the participant with the presence of an experimenter when trying to understand the material.

We defined four tasks: two debugging tasks and two type hierarchy tasks asking for details of a given domain (e.g. "Can a Spleus be married to a Schlok?"). As a within subjects design was considered most suited (to control for individual differences within such a small sample), it was necessary to construct two tasks (matched in difficulty) for each of these two types to compare the effects of having the tools available. The two tasks to test syntax highlighting presented the user with domains that were 54 lines in length, consisted of 1605 characters and contained 17 errors each. Errors were distributed evenly throughout the domains and were categorized into different types. The occurrence frequencies of these types were matched across domains as well, to ensure equal difficulty for both domains. To test the type diagram generator, two fictional domains with equally complex type hierarchies consisting of non-words were designed (five and six layers in depth, 20 and 21 types). The domains were also matched in length and overall complexity (five and six predicates with approximately the same distribution of arities, one action with four predicates in the precondition and two and three predicates in the effect).

Each participant started either with a debugging or type hierarchy task and was given the myPddl tools either in the first two tasks or the second two tasks, so that each participant completed each task type once with and once without myPddl. This results in 2 (first task is debugging or hierarchy) $\times$ 2 (task variations for debugging

[6] In Usability Engineering, a typical number of participants for user tests is five to ten. Studies have shown that even such small sample sizes identify about 80 % of the usability problems [11, 12]. Our study design required at least eight participants.

and hierarchy) $\times$ 2 (starting with or without myPddl) = 8 individual task orders, one per participant.

For the debugging tasks, participants were given six minutes[7] to detect as many of the errors as possible. They were asked to record each error in a table (pen and paper) with the line number and a short comment and to immediately correct the errors in the code if they knew how to, but not to dwell on the correction otherwise. For the type hierarchy task, participants were asked to answer five questions concerning the domains, all of which could be facilitated with the type diagram generator, but one of which also required looking into the code. Participants were told that they should not feel pressured to answer quickly, but to not waste time either. Also they were asked to say their answer out loud as soon as it became evident to them. They were not told that the time it took them to come up with an answer was recorded, since this could have made them feel pressured and thus led to more false answers. At the end of the usability test they were asked to evaluate the perceived usability of myPddl using the system usability scale [13].

4.2.2 *Results*

- Debugging Tasks

  As shown in Figure 3, on average participants found 7.6 errors without syntax highlighting and 10.3 errors with syntax highlighting (i.e. approximately 36 % more errors were found with syntax highlighting). Two participants remarked that the syntax highlighting colors confused them and that they found them more distracting than helpful. One of them mentioned that the contrast of the colors used was so low that they were hard for her to distinguish. She found the same number of errors with and without syntax highlighting. The other of the two was the only participant who found less errors with syntax highlighting than without it. With myPddl-syntax, two participants found all errors in the domain, while none achieved this without syntax highlighting.

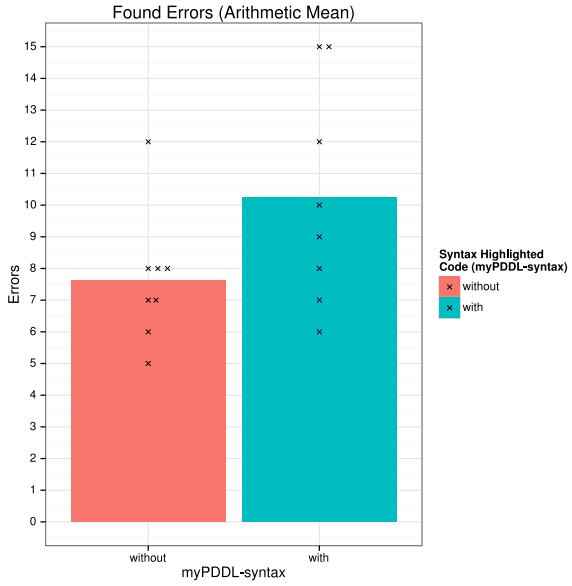

Figure 3: Comparison of detected errors with and without the syntax highlighting feature. Each cross ($\times$) shows the data value of one participant. The bars display the arithmetic mean.

- Type Hierarchy Tasks

[7] A reasonable time frame tested on two pilot tests.

Figure 4 shows the geometric mean[8] of the completion time of successful tasks for each question with and without the type diagram generator. With the type diagram generator participants answered all questions (except Question 4) on average nearly twice as fast. The fact that the availability of tools did not have a positive effect on task completion times for Question 4 can probably be attributed to the complexity of this question. In contrast to the other four questions, to answer Question 4 correctly, the participants were required to look at the actions in the domain file in addition to the type diagram. Most participants were confused by this, because they had assumed that once having the type diagram available, it alone would suffice to answer all questions. This initial confusion cost some time, thus negatively influencing the time on the task.

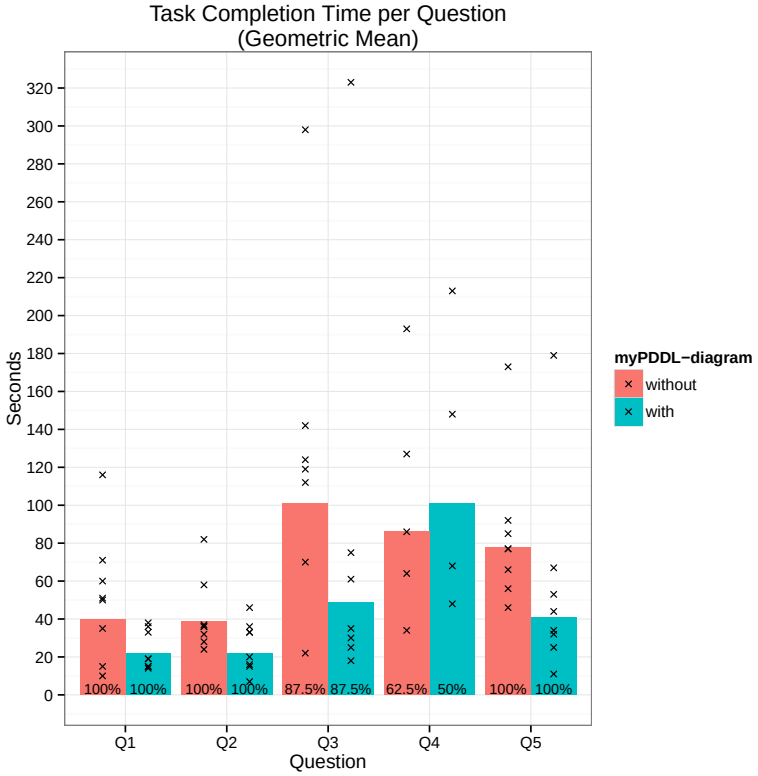

Figure 4: Task completion time for the type hierarchy tasks. The bars display the geometric mean. The percent values at the bottom of the bars show the the percentage of users that completed the task successfully.

- System Usability Scale

  myPddl reached a score of 89.6 on the system usability scale 10 , with a standard deviation of 3.9. Since the overall mean score of the system usability scale has an approximate value of 68 with a standard deviation of 12.5 [15], the score of myPddl is well above average with a small standard deviation. A score of 89.6 is usually attributed to superior products [16]. Furthermore, 89.6 corresponds approximately to a percentile rank of 99.8 %, meaning that it has a better perceived ease-of-use than 99.8 % of the products in the database used by Sauro [15].

[8] The geometric mean is a more accurate measure of the mean for small sample sizes as task times have a strong tendency to be positively skewed [14].

MYPDDL was designed with the goal to support plan engineers in modeling domains and planning problems as well as in understanding, modifying, extending, and using existing planning domains. This was realized with a set of tools comprising code editing features, namely syntax highlighting and code snippets, a type diagram generator, and a distance calculator. To also have all tools accessible from one place, they were made available in the Sublime Text editor. The different needs and requirements of knowledge engineers are met by the modular, extensible, and customizable architecture of the toolkit and Sublime Text. The evaluation of MYPDDL has shown evidence that it allows a faster understanding of the domain structure, which could be beneficial for the maintenance and application of existing task specifications and for the communication between engineers. Users perceive it as easy and enjoyable to use, and the increase in their performance when using MYPDDL underpins their subjective impressions. Despite MYPDDL already providing a rich modeling environment, there are still numerous features that could be added in the future. Especially MYPDDL-CLOJURE offers multiple interesting further research directions: It provides a basis for dynamic planning scenarios. Applications could be the modeling of learning and forgetting (by adding facts to or retracting facts from a PDDL file) or the modeling of an ever changing real world via dynamic predicate lists. Another way of putting the interface to use would be by making the planning process more interactive, allowing for the online interception of planning software in order to account for the needs and wishes of the end user.

[1] Tomas Plch et al. "Inspect, edit and debug PDDL documents: Simply and efficiently with PDDL Studio". In: *ICAPS12 System Demonstration* (2012).

[2] Stefan Edelkamp and Jörg Hoffmann. "PDDL2.2: The language for the classical part of the 4th International Planning Competition". In: *4th International Planning Competition (IPC-04)* (2004).

[3] Tiago Stegun Vaquero, Flavio Tonidandel, and José Reinaldo Silva. "The itSIMPLE tool for modeling planning domains". In: *Proceedings of the First International Competition on Knowledge Engineering for AI Planning, Monterey, California, USA* (2005).

[4] Tiago Stegun Vaquero et al. "On the Use of UML.P for Modeling a Real Application as a Planning Problem". In: *ICAPS*. 2006, pp. 434–437.

[5] Flavio Tonidandel, Tiago Stegun Vaquero, and José Reinaldo Silva. "Reading PDDL, writing an object-oriented model". In: *Advances in Artificial Intelligence – IBERAMIA-SBIA 2006*. Springer, 2006, pp. 532–541.

[6] TS Vaquero et al. "itSIMPLE4.0: Enhancing the modeling experience of planning problems". In: *System Demonstration–Proceedings of the 22nd International Conference on Automated Planning & Scheduling (ICAPS-12)*. 2012.

[7] Rich Hickey. "The Clojure programming language". In: *Proceedings of the 2008 symposium on Dynamic languages*. ACM. 2008.

[8] Tim Lindholm et al. *The Java virtual machine specification*. Pearson Education, 2014.

[9] Simon Parkinson and Andrew P Longstaff. "Increasing the numeric expressiveness of the planning domain definition language". In: *Proceedings of The 30th Workshop of the UK Planning and Scheduling Special Interest Group (PlanSIG2012)*. UK Planning and Scheduling Special Interest Group. 2012.

[10] MMS Shah et al. "Knowledge engineering tools in planning: State-of-the-art and future challenges". In: *Knowledge Engineering for Planning and Scheduling* (2013).

[11] Jakob Nielsen. "Estimating the number of subjects needed for a thinking aloud test". In: *International journal of human-computer studies* 41.3 (1994), pp. 385–397.

[12] Wonil Hwang and Gavriel Salvendy. "Number of people required for usability evaluation: the 10±2 rule". In: *Communications of the ACM* 53.5 (2010), pp. 130–133.

[13] John Brooke. "SUS – A quick and dirty usability scale". In: *Usability evaluation in industry* 189 (1996).

[14] Jeff Sauro and James R Lewis. *Quantifying the user experience: Practical statistics for user research*. Elsevier, 2012.

[15] Jeff Sauro. *A practical guide to the system usability scale: Background, benchmarks & best practices*. Measuring Usability LLC, 2011.

[16] Aaron Bangor, Philip T Kortum, and James T Miller. "An empirical evaluation of the system usability scale". In: *Intl. Journal of Human–Computer Interaction* 24.6 (2008), pp. 574–594.

Feature	Function	PDDL STUDIO	ITSIMPLE	PDDL-mode	MYPDDL
latest supported PDDL version	considering recent PDDL features	1.2	3.1	2.2	3.1
syntax highlighting	supporting error detection and code navigation	yes	basic	basic	yes
semantic error detection	supporting error detection	yes	no	no	no
automatic indentation	supporting readability and navigation	no	no	yes	yes
code completion	speeding-up the knowledge engineering process	yes	no	yes	yes
code snippets	speeding-up the knowledge engineering process externalizing user's memory	no	yes	yes	yes
code folding	supporting keeping an overview of the code structure	yes	no	yes	yes
domain visualization	supporting fast understanding of the domain structure	no	planned	no	yes
project management	supporting keeping an overview of associated files	yes	yes	no	yes
UML to PDDL code translation	supporting initial modeling	no	yes	no	yes
planner integration	allowing for convenient planner access	basic	yes	no	basic
plan visualization	supporting understanding and crosschecking the plan	no	yes	no	no
dynamic analysis	supporting dynamic domain analysis	no	yes	no	no
declaration menu	supporting code navigation	no	no	yes	no
interface with programming language	automating tasks extending PDDL's modeling capabilities	no	no	no	yes
customization features	acknowledging individual needs and preferences	basic	no	yes	yes

Table 1: Comparison of knowledge engineering tools and their features.