**Frege IDE**

Master Thesis

*(Please note, this is a work in progress, the text and the graphical layout will be updated)*

**Outline**

* Chapter 1: Introduction
  + Motivation / Introduction *(what was already done, why this work even exists, how this relates to other IDEs, what is Frege)*
  + Goals *(What I strived to achieve in this work, what subset of features should make / made it here, why exactly this, how this relates to projectional IDEs – MPS, why choose a projectional IDE, what is MPS – very shortly)*
  + Outline *(what other chapters are here, what can be found there)*
* Chapter 2: Meta-programming system / JetBrains MPS
  + - *what it is, what it can do, links & references*
    - *difference between projectional and classic text-based IDEs, how this relates to MPS, what makes MPS different*
    - *examples (images), what it would be like to use MPS, so the reader has a clearer picture*
    - *working with MPS, structure, editor, constraints, behaviour, checking rules, typesystem, textgen – generic explanation + exemplar language (not necessarily tutorial, but the explanation should be clear and how it would be used)*
* Chapter 3: Frege
  + *what is Frege*
  + *Frege vs. Haskell*
  + *Frege language, examples, and what we decided to support in 'Frege-IDE'*
* Chapter 4: The grammar
  + *Official Frege grammar, or references (appendix?)*
    - *This is too abstract. Maybe pick only some interesting parts?e.g. “this is usually solved as...”, “on this level it means this and this is supported”...?*
  + *Subset related to what we decided to support in “Frege-IDE”*
    - *(Analysis, what parts of the grammar we decided to cut off, what couldn't be cut off, hot it relates to everything)*
* Chapter 5: Frege in MPS
  + *Grammar transformation for MPS structure aspect, design of the Frege structure, how and why (analysis + design decisions)*
  + *Editor*
    - *“Visuals” (FunctionName FunctionParameters = Expression)*
    - *Editor transformation and substitution actions*
    - *Lacking documentation on the MPS part.*
  + *References, scopes*
    - *(Constraints + behaviour)*
  + *Typesystem*
    - *Expression evaluation*
    - *TBA*
* *Chapter 6: Evaluation*
  + *What we strived to achieve, again? Recapitulation?*
  + *User-friendliness of the “Frege-IDE”, against classical text-based IDEs, advantages and disadvantages*
    - *(Should be probably mentioned user's time investment into learning to use the IDE)*
  + *Limitations*
    - *Something couldn't be done easily, e.g. Enter key-press does not always create a new line, ambiguity*
    - *Not all transformations are possible, e.g. rewriting f x y = x + y to f , x, y :: Int -> Int -> Int (by removing = and adding , , ::) is basically impossible to cover – it is not a text editor*
    - *Intentions – for the example above intention is an option to cover such transformation*
    - *(Mention built-in JAVA-like language, that also has these limitations)*
  + *Experience*
    - *What I found MPS lacking about, flexibility, not much detailed documentation*
* *Conclusion*
  + *Are projectional IDEs good for functional languages?*
  + *Is “Frege-IDE” usable?Future work, possible extensions.*
* *References*
* *Appendix*
  + *Frege formal grammar*
  + *Source code description*
  + *Examples*
  + *User manual for “Frege-IDE”*
* Chapter 1: Introduction
* Motivation / Introduction *(what was already done, why this work even exists, how this relates to other IDEs, what is Frege)*

(Classic intro,a lot of wiki references to say the obvious and well known.)

(what are IDES) (<https://en.wikipedia.org/wiki/Integrated_development_environment>) Integrated development environments (abbreviated IDEs) are a set of software applications that provide tools and facilities to computer programmers. They greatly ease the process of software development, providing features like intelligent code completion, syntax highlighting, build automation tools, debugger, etc. (…add anything??)

(what is Frege) (<https://en.wikipedia.org/wiki/Frege_(programming_language)> Frege, named after the German mathematician, ~~logician and philosopher~~ [Gottlob Frege](https://en.wikipedia.org/wiki/Gottlob_Frege), is a functional language heavily based on Haskell (add ref. link) for the Java virtual machine. It is considered a Haskell dialect, sometimes called “a Haskell for the JVM”. (…todo)

known IDEs

There are several IDEs for Haskell, not many specifically built to support Frege. Examples can be found at <https://wiki.haskell.org/IDEs>, most of them providing mainly syntax highlighting, macros and project management features, while some also provide more advanced features like code completion or type evaluation and inspection.

That being said, all of the notable IDEs are text-based, (add explanation?

See <https://en.wikipedia.org/wiki/Structure_editor>

…most source code editors are instead text editors with additional features such as syntax highlighting and code folding, rather than structure editors. The editors in some integrated development environments parse the source code and generate a parse tree, allowing the same analysis as by a structure editor, but the actual editing of the source code is generally done as raw text.

). In this work, however, we try to approach the topic from a different view and try to design a projectional editor (/IDE) specifically for the Frege programming language.

(what is projectional editor) quoting <https://en.wikipedia.org/wiki/Structure_editor>

A projectional editor (aka structure, structured editor), is a document editor that is cognizant of the document’s underlying structure. They are usually used to edit hierarchical or marked up text, computer programs, diagrams, and any other type of content with clear and well-defined structure. While for the most computer programs a conventional text-based IDE may be more suitable, for specific programming languages, especially DSL (<https://en.wikipedia.org/wiki/Domain-specific_language>) a projectional editor might prove to be a more effective tool.

<https://martinfowler.com/bliki/ProjectionalEditing.html> *- additional info sources*

In this work we want to design a projectional editor supporting a certain subset of the (features of) Frege programming language and examine what are the advantages and disadvantages of such approach over creating a conventional text-based IDE.

* Goals (What I strived to achieve in this work, what subset of features should make / made it here, why exactly this, how this relates to projectional IDEs – MPS, why choose a projectional IDE, what is MPS – very shortly)

As was already mentioned in this paper, in this work we design a projectional editor for a subset of (features of) Frege programming language.

We have chosen JetBrains® *(legality, trademarks, etc.? check)* MPS platform as an underlying tool for designing our IDE. (briefly - why:) MPS (standing for “Meta-programming system”) is an open source software solution allowing developers and language designers to create different kinds of projectional IDEs. It is a so called language workbench (A language workbench is a software development tool designed to define, reuse and compose domain-specific languages together with their integrated development environment. - <https://en.wikipedia.org/wiki/Language_workbench>)

It allows for creating both simple and complex languages, especially DSLs, while also allowing extending existing ones, when the languages available do not meet the needs of a developer. *(quoting MPSLW I.)* MPS has a large set of features allowing for designing editors which closely resemble those from conventional, text-based, IDEs.

Frege, based on Haskell language, has rather many syntactic (and semantic) constructs for this work to be able to include them all. We have therefore focused our attention only on the most important features worth examining, such as function declaration and definition, operators and (custom) datatypes. Our ideal IDE will have a user-friendly editor that should emulate normal text editing and writing code in the way that most Frege and Haskell developers are used to. This should be accompanied by a context help, sometimes referred to as ‘intellisense’, which would allow for referencing already defined functions, operators, variables, etc., in the corrects spots in the code. Last, but not least, we will strive for a type checker, which would be able to find small mistakes in the code, such as calling a function with illegal arguments, or evaluate type of an expression.

* *Outline (what other chapters are here, what can be found there)*

More about MPS tool is written in chapter 2, which examines what MPS is, what it can do and what are its limitations. The chapter examines a project structure in MPS, how to define an editor for a simple language and how to tackle certain problems.

Frege is described more in chapter 3, where we also take a look into what features we actually wanted to implement in our IDE.

Chapter 4 examines Frege grammar and how it relates to the features we want to support. We describe how the grammar should be transformed into concepts in MPS tool on the high-level. We provide a brief analysis and try to explain our decisions.

Chapter 5 is dedicated to the concrete work implementation. We show how exactly the grammar from chapter 4 was transformed into MPS concepts, how the editor was designed, how we tackled problems with references and context help. We conclude the chapter with type system, where we describe some of the more interesting algorithms used in the work, such as the one to evaluate type of expressions or inference types of function arguments, when provided with the function’s annotation.

We evaluate our decisions in chapter 6 where we also examine the advantages and disadvantages of the implemented projectional editor over standard text-based IDEs.

A brief summary of the whole work may be found in the conclusion, where we strive to answer the final question, whether projectional IDEs are good for functional languages.

After that, references of the used literature may be found and appendix, which provides some of the more interesting source materials.

* Chapter 2: Meta-programming system / JetBrains MPS
  + - *what it is, what it can do, links & references, difference between projectional and classic text-based IDEs, how this relates to MPS, what makes MPS different*

<https://www.jetbrains.com/mps/>, <https://www.youtube.com/watch?v=XGm_khXZl44&feature=youtu.be>

Meta-programming system, or MPS, is an open-source language workbench that focuses on DSL (domain-specific languages). [JB-YT] It is a software engineering tool which helps its users to create a new language and then to write other programs in that language. [FC] MPS has a wide range of users, with areas currently being applied in including electrical engineering, data mining, insurance industry and other. The tool can be used to create new languages as well as extending existing ones. (It should be noted that only the languages created in MPS tool may be further extended.) Programs written in the defined languages may then be conveniently transformed into pure text in a specific, usually generic-purpose language. [JB-YT]

MPS is a projectional editor, which means it does not treat the document as a text, but rather as structured concepts. This allows its users to create languages which involve non-parsable notations, such as decision tables, diagrams, and other controls. [JB-YT] Additionally, several editors may be specified for a single language, thus allowing users to switch between different visual representations of a document.



2.1 View of ‘math’ demo language in MPS which includes non-text elements to extend the existing Java language

Traditional process of compiling written code involves lexers and parsers to read programs, which are expressed as text files. The two combined then transform the code into abstract syntax tree (AST) data structures [FC], from which in the process of semantic analysis and code generation an executable program is created. A similar technique is used in the text-based IDEs. In contrast, in MPS, the user works with AST directly, therefore completely omitting the process of lexical analysis and subsequent parsing. [FC] This brings certain advantages:

* It may be easier to extend an existing language.

Extending lexers and parsers to accommodate for the changes in a language requires a certain set of skills and a deeper knowledge of the language’s grammar. The process is complicated, since it requires a programmer to keep track of the possible ambiguities that may arise when defining new grammar rules for a parser.[FC] (See, so called, ‘Dangling else’)

On the other hand, in MPS, the process usually only requires defining new concepts that can act as AST nodes and specifying places in the corresponding AST, where the new nodes can be created.

This also means that in MPS we can combine syntax of several different languages and introduce no syntax ambiguities whatsoever. (This however still may look ambiguous to the user, if there are several different concepts with the same textual representation.)

* We can check for type errors and other mistakes in the code at almost any time.

When writing a code in a traditional environment, one has to define a specific set of rules to deal with the incorrect syntax. Code being currently written means, it almost certainly cannot be correctly evaluated by the defined parser, so in an example case such as below, we will not be able to tell the user right away that the integral and string types are incomparable between themselves:

f = if 1 = "hello" then |

(syntax error on '=')

This cannot happen in MPS, since the code is already ‘parsed’. Even though ‘then’ and ‘else’ branches are not set yet, we can already check for errors in ‘condition’ node (1 = "hello"). This is also useful when designing a smart code completion feature, which requires a certain knowledge of the surrounding code context. This is impossible if we cannot parse the code beforehand, unless certain ‘tricks’ are used.

****

2.2 Exemplar AST for a simple arithmetic expression

Since a user may in MPS work only with AST directly, several features are supported to allow easy transformations of ASTs and allow creating IDEs that may even closely resemble the traditional, text-based, ones. What this means is that we can define actions how the AST should transform when, for instance, a certain node (or a whole subtree) is deleted, a certain specific text is written at the end or a beginning of a visual representation of a node, etc. The MPS actions are described further in this chapter.

* *working with MPS, structure, editor, constraints, behaviour, checking rules, typesystem, textgen – generic explanation + exemplar language (not necessarily tutorial, but the explanation should be clear and how it would be used)*

A project in MPS is divided into two main categories: solutions and languages.

A ‘language’ is the user defined language. It may represent a completely new language or an extension of an existing one. There may be several different languages defined in a single MPS project. They can act as an extension of each other or be completely independent languages.

A ‘solution’, on the other hand, is a part of the project that represents documents, or a code, written in (usually) one or more of the defined languages. Sometimes the solution only acts as a runtime support for one (or more) of the defined languages to be used, for example, in a code generation process.

******

2.3 Logical view of a project in MPS

Solution is a set of models. They act as packaging units that make it possible to reference the corresponding set of models from other solutions of languages. [FC]

Model is simply a set of ASTs. In the languages used by a solution, a certain AST nodes may be set to act as a root node. Model then basically consists of such AST root nodes. Each AST is basically analogous to a single source file in traditional programming paradigm. [FC]

Language describes what types of ASTs can be created with the language. [FC] Furthermore, it includes visual representation of each node, AST transformation actions, syntax and semantic rules and many other ‘settings’. It is separated into several categories, named aspects. We will describe only the most important ones that were also used in this work.

**Structure**

<https://confluence.jetbrains.com/display/MPSD20173/Basic+notions>

Structure aspect of a language allows to define structure of possible ASTs that can be expressed with the corresponding language. [FC] It defines what kind of nodes may be used in user models, what properties, children and references they may have. [JB-D]

Key notion in this aspect is a ‘concept’. Concept represents a sort of a class of AST nodes. It specifies what children and properties all nodes of the ‘class’ have, or can have. It closely resembles working with classes and instances in many popular object-oriented programming languages. In this analogy, a concept is basically a class, which extends another class. (All concepts have to extend at least BaseConcept, similar to Java language, where each class extends Object class.) It may be defined as abstract, in which case no AST nodes may be created directly for such concept. We may define interfaces, which concepts may implement, concept properties, which are analogous to fields and attributes of primitive types (this is not exactly true as property may be of type string or enumeration), children, which represent public fields of object types (i.e. referencing an AST node of a certain type with ownership, not necessarily exclusive) and references, which are similar to children but without the ‘ownership’ part.

Each concept may specify visual appearance of the node (more in ‘editor aspect’), implement methods (behavior aspect) and specify additional syntax constraints on where they can actually be placed in AST (constraints aspect).

******

2.4 Definition of a new concept in MPS

We start by naming a concept. This is similar to naming a class in languages like Java and must follow a similar set of naming rules.

The ‘extends’ clause provides a reference to the super-concept. By default, concepts are created with BaseConcept as their super concept, but this can be changed to a more specific super-concept. This basically encodes ‘is-a’ relationship in UML. [FC] If concept A extends concept B, it indicates that concept A has all of properties, children, references, methods, and definitions from all other language aspects, as B.

Concept may implement interface(s) by using ‘implements’ clause. More about interfaces is discussed later in ‘structure’ chapter.

A concept may be set for its instances to act as root AST nodes. As it was mentioned before, a solution is a set of models all of which consist of root AST nodes. Only the concepts denoted as such may be used that way.

Alias acts as string that triggers a built-in auto-completion menu. If the name is unambiguous (i.e. not a prefix of another item in the menu), an instance of the concept is immediately created. More about menu is discussed in editor aspect of language.



2.5 Auto-completion menu in MPS

Properties hold values that are owned by the concept. They are analogous to UML class properties. [FC] They are either:

* Primitive type – integer, boolean, or string
* Enumeration type – a custom ‘enum data type’ may be created in structure aspect
* Constrained data type - a custom ‘constrained data type’ may be created in structure aspect, which is basically a string compared to a defined matching regular expression

It may be interesting to note that primitive types can be derived from using the other two options.

Children resemble aggregation relationship known from UML. They are other concepts with defined cardinality ([1], [0..1], [1..\*] or [0..\*]) and type. A concrete instance (i.e. a node) must be a child of an exactly 1 other node, with an exception of root nodes. This way their life depends on their parent’s life meaning that removal of a certain node from AST removes all of its children as well as children of their children, recursively down to leaf nodes.

Expressing a relationship between nodes can be also done via references. It is only possible to create a reference to a node if this node already exists in the corresponding AST. [FC] Contrary to children, cardinality can take here only two forms: [0..1] meaning an optional reference and [1] meaning a required reference to a node of the given type.

Where would a reference be used? Consider a following piece of code in Frege:

f = 7

g = 1 + f

We could express the formed AST in many different ways, but let us imagine for the sake of simplicity a root node consisting of statement nodes. Both f = 7 and g = 1 + f are statements. It is easy to imagine 1 + 2 as a tree with node + having 2 children: “1” and “2”. But in the case of 1 + f it is less clear what f is. Using a reference here might be helpful. We already have a statement declaring, what f is, in the AST. In 1 + f we are only applying an existing function f. Such AST might then look like this:



2.6 Simplified AST for statements with a reference

Interface concepts are a mechanism to declare concept characteristics that can be used across several concept types. [FC] Unlike concepts, we cannot define alias for them nor can they extend concepts, only other interfaces. They are mostly used for grouping properties that are commonly used together and passing them onto necessary concepts. [FC]

**Editor**

[**https://confluence.jetbrains.com/display/MPSD33/Editor**](https://confluence.jetbrains.com/display/MPSD33/Editor)

Editor aspect is responsible for rendering and editing ASTs by the user of the language being created. This includes textual and graphical representation of each AST node and certain AST transformation actions. This aspect is what makes MPS a projectional editor, rather than using lexers and parsers to process the user-written code.

****

2.7 Concept editor in MPS

The easiest way to define an editor for a language is to define an editor for each concept (called ‘concept editor’). There may be several different editors defined for a single concept, which offers different views of the same concept for different needs. [FC] If no editor is defined for a given concept, a default one will be provided by MPS.

On the other hand, editor component is an editor responsible for rendering and editing only a part of a node. [FC] It does not focus on any single concept and as such may be reused across several concept editors to render certain parts similarly. [FC]

As most aspects in MPS, editor is too defined by writing a code in a special language provided by MPS – in this case jebtrains.mps.lang.editor. Furthermore, the code is written inside a projectional editor – to define a specific projectional editor for the new language being created.

The editor language is used to specify editor cells. An editor cell is typically used to render a certain portion of a node over a rectangular region in MPS editor window. The types of editor cells include:

* Constant cells – These are used to render keywords and other constant text in editor. [FC]
* Property cells – They render content of a specific property of a concept for which the editor is being defined. Editing such a cell in the editor window for a concrete AST is immediately reflected in the given property of the corresponding AST node. The cell provides automatic binding to the concept’s property.
* Child cells – Child cells delegate the rendering of a specific concept’s child (or a set of children) to their corresponding concept editors. The concrete behavior of such a cell depends on the child’s cardinality:
  + [1] – the editor cell is always present
  + [0..1], [0..\*], [1..n] – child nodes are bound to their corresponding editors and removing a child in MPS editor window results in removing it from the parent node of the corresponding AST as well
  + [0..\*], [1..\*] – the children’s corresponding concept editors are separated by a specified (textual) delimiter
* Referent cells – These cells are used to display an attribute of a referenced node from the given concept (see references in “structure” above). As in case of property cells, they are mapped to a certain property of the referenced node in the AST. However, they can only reflect the property of the original node, but not affect it.

******

2.8 Exemplar usage of the editor defined for Java-like language - changes in the name property of the original node (left) result in immediate re-rendering of the referent cell (right)

* Collection cells – Wrapper-like cells to contain other editor cells. They affect visual arrangement of the cells being rendered. There are 3 types:
  + Horizontal cells
  + Vertical cells
  + Indent cells
* Model access cell – This type of cell is used to modify a state of a node(s). We can define custom getter and setter for the cell, or just a custom getter, to change the behavior of our editor.

There are several other types of editor cells. Here we only described the most-used ones.

To change the way an editor cell is rendered, we can use editor styles. Applying editor style could be described as analogous to applying CSS *(Cascading Style Sheets)* styles to DOM nodes in HTML and XML documents. This allows us to change colors of a text, background color, spacing, padding, functional aspects such as editor cell being editable, read-only, etc.

******

2.9 Editor style for a selected editor cell

While all of this allows us to create a visually appealing editor, it still may not come as a user-friendly, especially to developers used to program in text-based IDEs. To simulate many types of behavior of classic text editors, MPS comes with a notion of actions. These were originally part of Action aspect of MPS, but since they are tightly related to the editor, they were moved in newer versions of MPS into this aspect.

To better explain the need of editor actions, consider an exemplar expression from figure (image?) 2.2: (7 + 1) \* 2 + 3. We already know what an AST would look like for such an expression, however, in MPS, without actions, a user would need to specify the AST him or herself.

First, we would need to add addition binary expression node, then for the right operand, literal ‘3’ while for the left operand another binary expression - multiplication, for which we would need to specify the operands again, and so on.

With a careful implementation of actions, however, we could let the user type the expression from ‘left to right’ while transforming the AST automatically for him. To simplify, let us forget about brackets for a while and try to type a slightly different arithmetic expression: 7 – 1 \* 2 + 3.

|  |  |
| --- | --- |
| First, the user types in ‘7’. That is a very simple unary expression and no further work is to be done here. | 2.10 |
| Then, user hits ‘-’. We immediately create a binary expression subtree, where root is the operator ‘-‘. ‘7’ is put as its left operand and we set focus on the right operand, so the user may edit that. | 2.11 |
| User types in ‘1’, which only concludes the editing of the right operand and we are done there. | 2.12 |
| Then, however, follows ‘\*’. We are now editing the right child of AST of the expression “7 – 1”. So we take a look at the parent’s operator’s precedence: - is definitely less precedent than \*. So we create a subtree for binary operator \*, put 1 as its left child and set focus on the right child. The subtree is placed in 1’s stead (i.e. right child of AST of the expression “7 – 1”). | 2.13 |
| Then comes 2, which only finishes editing of the right child of \* subtree. The twist comes, however, when user types in ‘+’. We take a look on the parent’s operator, which is \*. This has a higher precedence, so a new subtree has to be created elsewhere now. The parent of the \* node is, however, -. While – has the same precedence, as +, all of the operators are left associative, which means, we have to create our new subtree even on the higher level. We create the subtree, put the current AST for the expression “7 – 1 \* 2” as its left child and set focus on its right child. Typing ‘3’ only finishes editing of the right child and we are done. | 2.14 |

In the description above, we were always editing a specific editor cell corresponding to a single node in the AST. We always handled an event of writing a specific textual pattern right of a certain editor cell. This is just what MPS allows us to do.

**Transformation menu**

The transformation menu provides a way to describe side-transformations and substitute actions in MPS. For a particular cell in the editor we can specify one of the two mentioned actions.

Side-transformations allow to manipulate the AST when a certain textual pattern is enterd either from the left or right of the current editor cell.

Substitute actions, on the other hand, are triggered, when one node is being substituted by another node. They allow to define a default concept to be used for instantiating new nodes in certain places in the model, or an action to take upon creating a new node of a certain type in certain places, and others. We descibre substitute actions later in this chapter.

<https://confluence.jetbrains.com/display/MPSD20172/Transformation+Menu+Language>

Transformation menu is always created for a specific set of concepts. Similarly to OOP languages, we can create a transformation menu for a concept, which will be also applicable to all of its (either direct or indirect) subconcepts. There are two main types: default and named menus.

Every concept has associated a default transformation menu. If the language designer does not provide one explicitly, a transformation menu defined for the closest super-concept is assumed. [JB-C] If none are defined, a one implicitly defined for BaseConcept is used. A named menu is an additional menu for a concept. Unlike the default menu, it is not applied to a concept directly, but rather has to be attached to a certain editor cell(s) of the given concept manually. Another way to use a named transformation menu is to include it into a different menu via „Include menu“ statement.



2.15 A default transformation menu for a concept

We start by specifying section, i.e. where the transformation should take place. Side transformations have only two options: either they are triggered upon typing a text right of an editor cell, or left of a cell.

Then, we specify a, so called, menu part [JB-C]:

* Action - a simple menu item specifying an action to be performed, its corresponding menu text and applicability. The 3 categories are separated (see figure 2.15).
  + Text – a string that triggers the current action written either left or right of an associated editor cell (depending on the „section“ specification). It can either be a constant or a piece of code that returns a string that triggers the action.
  + Can execute – a piece of code that is executed upon the text trigger. If the code returns false, the current action may not be executed. *(Returning „true“, however, does not indicate that the action should just immediately execute. If, for example, we have another action for the same concept defined with different „Text“ trigger, for which its „can execute“ clause would also return true, ambiguity arises. The „Text“ part is just a trigger to execute any of the actions for which „can execute“ returns true. That is why it is mostly recommended to leave* “*can execute <always>“, which has a different behaviour – the action is triggered exactly when the string from „text“ part is typed in.)*
  + Execute – a code associated with the current action, i.e. what should actually happen.
* Group - a collection of menu items.
* Include ... - include a specific default or named menu. We may also include a transformation menu not aimed for the concept we are currently dealing with. In that case we have to provide the node of that specific type (e.g. be it a parent of a current node, or a completely new node created ad-hoc in the AST, etc.)

There are also other menu parts in MPS we will not be currently dealing with.

Important thing to note when dealing with transformation menus is that they only apply on constant, property and referent editor cells. If a concept has other children for which it defines child editor cells in the concept editor, then they do not trigger the concept’s side transformations. In that case a language designer has to pay a greater attention to what cells actually trigger which action.



2.16 In Frege, we may define custom algebraic datatypes as follows:

data Maybe a = Just a | Nothing

We call “Just a” and “Nothing” datatype constructors. After each one it is desirable to be able to add a new constructor when typing character ‘|’. But since a constructor can (but does not have to) be followed by one or more type variables, like ‘a’ in ‘Just a’, it is not sufficient to specify the transformation for the ‘DataConstructor’ concept only. Even more troubling, the transformation right of a type variable is desirable only if it happens to be the ‘right-most’ editor cell in the given ‘DataConstructor’. (i.e. we would not want to add a new constructor when typing ‘|’ right of ‘a’ in this case: data Maybe a b = Just a| b ) The solution is to simply include the default transformation for ‘DataConstructor’ into type variable (here: ‘Simpletype’) concept, but only if it happens to be the right-most child of a ‘DataConstructor’.

Substitute actions differ from side transformations. Typically, they are actions from completion menu, which is invoked in MPS by pressing ctrl + space. Since MPS is not a classic text editor, completion menu is a vital part of the tool and a user will find him or herself using it quite often.

A default substitute menu is always provided by MPS when the caret‘s position is in front of a node or the node is selected. *(We mean here a group of editor cells corresponding to a single AST node.)*



2.17 Invoking a substitute menu on a selected node

For that case a menu with all concepts applicable in the given context are displayed, which can substitute the current node. [JB-C] Abstract concepts are, understandably, not displayed as well as none of the concepts, for which their constraints do not allow their presence in the current place. (See constraints aspect for more information). Smart references are treated somewhat otherwise. A smart reference is a term we use for concepts that contain only a single reference and nothing else. Such a concept is not added to the completion menu itself, but rather all of the nodes it could contain a reference to *(i.e. the nodes in the current scope)* are added to the completion menu. Selecting such a node results in creating an instance of the corresponding smart reference with the reference to the selected node.

Similarly to transformation menus, substitute menus are tied to a concept type and can be either default, or named. Named substitute menus can, however, be only included in other menus and on their own have no effect on the editor.

Adding an empty default substitute menu for a concept makes it inelligibile to be included in the default menu provided by MPS. This is a useful way to treat concepts like ‚empty line statement‘. In such an example, we would set that a default ‚statement’ is ‚empty line‘ (in ‚element factory‘ settings for ‚child editor cells‘ in ‚concept editor‘). But since it does not make much sense to replace a statement with an empty line (usually a user would want either to delete the node completely or add a new line above), we will remove it from the completion menu this way.

There are several entries to specify for a substitute menu:

* Include – includes a specified substitution menu.
* Subconcepts menu - includes all subconcepts of the current concept.
* Concept list - adds a specified collection of concepts.
* Substitute action - adds a single substitute action.
* Wrap substitute menu - wraps a specified concept using the provided handler. [JB-C]

There are several more entries to choose from, but for the most common cases these ones are sufficient. We will now take a closer look onto entries „substitute action“ and „wrap substitute menu“, since they are a little more complicated.

Substitute actions allow to define an action to take upon writing a specific textual pattern in place of a node for which the substitute action is being defined. It consists of 3 main parts:

* Create node – a piece of code which returns a node of such a type that the current concept can be substituted for.
* Matching text – a string (or a piece of code which returns a string) that triggers the current substitute action.
* Can substitute – a boolean depending on whether the substitute action can execute. *(Similar paradigm as in „Transformation action“ applies here – returning true for several substitute actions with different „matching text“ still results in ambiguity and the action will not execute on its own, but will rather require user’s decision from the completion menu. If an additional check besides „matching text“ is necessary, it is recommended to check for user-entered string here as well. Otherwise it is best to omit this part completely, which will result in the action triggering only upon exact match of the user-entered text with „matching text“)*

For better undestanding, let us assume an abstract concept „Literal“ with two subconcepts: „Integer“ and „Boolean“. We want to automatically create a node, where a node of type „Literal“ is expected, upon user typing either an integer number (we create an instance of „Integer“ concept) or „true“ or „false“ (we create an instance of „Boolean“ concept). This is done by defining a default substitute menu for concept „Literal“ with 2 substitute actions. In both cases, „matching text“ simply returns any text a user has typed, since there is no single string, that would suit. Instead, we have to check, whether the written string is actually an integer number or „true“/“false“ in „can substitute“ clause. Finally, „create node“ clause simply handles the substituion action.

„Wrap substitute menu“ has a similar behaviour as smart references. In this case, however, a language designer specifies a concrete concept of which node instances should be suggested in the completion menu to substitute the current concept for. This also works for an example, where we have already defined substitute actions for literals, such as numbers, or boolean values. Let us assume such literals inherit from common ancestor – concept „Literal“. However, consider that we are defining a node of type „Pattern“. Now, „Pattern“ is not an ancestor for „Literal“ in terms of inheritance. This means that if we are to allow specifying literals where concept „Pattern“ is expected, we have to create a new concept, a „wrapper“ for concept „Literal“, that inherits from concept „Pattern“. (This is analogous to adapter design pattern.) Let us call it „PLiteral“. Now the problem is, the substitute actions defined for „Literal“ concept do not work here, since „PLiteral“ is expected instead of „Literal“. A simple solution is to define „wrap substitute menu“, which acts as a substitute menu for „Literal“, but as its output a node of type „PLiteral“ is produced instead.



2.18 Wrap substitute menu for concept PLiteral, which wraps concept Literal

From editor actions we mention also Cell Action Map. These are custom defined actions to take upon a user attempt to e.g. remove an editor cell, insert a new one, pressing a backspace key, etc. They are always associated with an editor cell (though several editor cells can share the same cell action map) and a simple event, such as cell removal, insertion, selection, copying and pasting.



* 1. Example of cell action map

A cell action map is always defined for a specific concept, but can be used for any editor cell in the corresponding editor of the concept. We then select an event we want to create a handler for. There are around 40 events to choose from, but probably the most important ones are INSERT and DELETE. ‘Can execute’ clause allows to specify a condition to cases where the action is actually applicable. Finally, in ‘execute’ we specify the handler itself. An example of where a ‘cell action map’ could be useful may be concepts which are semantically different, but syntactically similar. For instance, consider in Frege an expression surrounded by brackets and a tuple:

(let x, y, z be functions, e.g. x, y, z :: Int -> Int)

f = (x) 1 -- Expression surrounded by brackets

g = (x, y, z) -- Tuple

While similar, the difference is significant – an expression surrounded by brackets may be used as a ‘function application’, i.e. we may specify arguments right of the expression. On the other hand, such a syntax would be illegal for tuples. This means we only need to allow specifying arguments, if the node is an instance of the concept ‘BracketedExpression’, but not for ‘Tuple’. However, we still want to allow a user to remove all but the last item of a tuple. Since tuple requires at least 2 items, we may do this by automatically converting it into ‘BracketedExpression’ once the second item has been deleted.

To conclude, editor actions provide a flexible way to build a user-friendly editor that mimics many features of a traditional, text-based, editor. However, understandably, it is impossible to allow the completely same behavior, since a user is editing the AST and not the text. This means that almost every editing feature has to be implemented manually. We can, however, optimize the editor for the most common cases at least.

**Behavior**

Behavior aspect allows to, simply said, define methods on concepts. If we take the analogy with OOP further, then structure aspect allows us to declare classes and their fields while behavior aspect allows us to declare and implement their methods, including constructors.

Constructor is a block of code which is executed when a new node of the concept is created. Certain exceptions exist, however, when the constructor is not run. These mainly include creating an instance of a node by using “new node<MyConcept>()” statement (e.g. in editor actions) instead of “new initialized node<MyConcept>()”.

Similar concept of methods as in languages like Java is present here. A concept may be associated with several methods with strictly defined visibility (public, protected, private). Methods that can be overridden in subconcepts have to be marked as virtual. Static methods are not attached to an instance of the concept but rather have to be called on the concept itself. (concept/MyConcept/.staticMethodName()) [FC]

Important characteristic of the behavior aspect is that it allows to traverse the AST being created. We can easily inspect parent and children of any node as well as nodes’ references (see structure –

concept references).

2.20 Behavior aspect in MPS

**Intentions**

These are special user interface elements that allow executing predefined actions in certain places. They usually perform some modification of the current AST. The intentions menu is usually invoked by pressing “alt+enter” and selection is confirmed by pressing “enter” key.

 2.21 Intentions in Frege-IDE example

Standard intentions are defined for a concrete concept. There are also “surround-with” types of intentions, which are similar to the regular ones, but are designed to surround a part of the AST with another construct. [FC]

In regular intentions, we also specify, whether the intention can be executed inside of any of the concept’s children. This allows for example to invoke an intention defined for “DataConstructor” concept even if the caret’s current position is over one of its type variables, e.g.:

data Maybe a = Just a|

Where | denotes the caret’s current position – which is on DataConstructor’s child, ‘TypeVariable’ concept. Next, we specify a textual description of the intention that will be shown to the user upon invoking the intentions menu. “isApplicable” method specifies, whether an intention is applicable for the current state of the AST. If the method returns false, the intention will not be shown in the menu. Finally, “execute” method specifies the intention’s action itself.

 2.22 Example of a concept intention definition

**Constraints**

This aspect lets a language designer to declare constraints that help him or her control where nodes of a language are allowed. [FC] We define constraints always for a specific concept. They allow specifying several attributes each of which has a different meaning.

2.23 Example of a constraint definition which puts restrictions on a referenced node “module”

“Can be child” allows defining a boolean method which returns, whether a node of the current concept can be a child in a specific AST context. [FC] If the method returns false, then such a node will not be suggested in the auto-completion menu. Attributes “Can be parent” and “Can be ancestor” work in a similar way.

Property constraint allows restricting a set of values of a concept property. Additionally, it is possible to define a getter and a setter for the given property.

Referent constraint makes it possible to restrict how references are established to nodes of the concept. [FC] This basically allows restricting what nodes will be “referentiable” from the given concept. Part “Scope” specifies scope of “referentiable” nodes. Part “Presentation” specifies how these nodes will be displayed in the auto-completion menu *(their textual representation)*. Scope is an object which defines a list of potential targets that can be referenced. It also helps to locate a suitable target from the given list based on what the user has entered in the corresponding place in the editor. When no scope is provided in the concept’s constraint aspect, all nodes of the appropriate type are considered eligible. The auto-completion menu is filled with nodes based on this rule and the scope, if specified.

2.24 An example of applicable nodes in the function-definition context of Frege-IDE. According to the given scope, variable “x” is part of a different function pattern definition and as such may not be suggested in the auto-completion menu for the definition of function “gg”.

It is important to note that part of the job for a Scope object is to return a reduced list of nodes that conform to the already entered string by a user. Consider an example from figure 2.24. If we were to enter a string “f”, a list would be reduced to contain only item “ff” as far as “referentiable” nodes are concerned, since “gg” and “y” *(and operators)* does not contain “f” as its substring. *(Note that items like “false” and “if” are not nodes to be referenced, but rather aliases of certain concepts, which we may then create instances from, too.)*

A language designer may specify his or her own implementation of “Scope”. The new class must inherit from the “Scope” class provided by MPS implicitly and implement the necessary abstract methods. Notable methods from the class include:

* **public abstract sequence<node<>> getAvailableElements(@Nullable() string prefix);** Returns all of the nodes from the scope that begin with a string “prefix”.
* **public abstract node<> resolve(node<> contextNode, @NotNull() string refText);** Returns a node, if the entered string “refText” can unambiguously determine a “referentiable” node from the current scope. (In 2.24 “ff” would be an example of such a string.)

Another point to note is that it is not certain what string is represented a certain node by. That is why the method **resolve** cannot work automatically. Additionally, the string defined in Scope object does not have to be equal to the one specified in “Presentation” part of a link in constraint aspect.

**Typesystem**

Typesystem aspect makes it possible to report semantic errors to the user of the language. [FC] On its highest level, it could be said it contains mechanisms to check for both non-typesystem and typesystem rules.

Non-typesystem rules are called checking rules. These serve a language designer to implement custom semantic error checks. An example in Frege-IDE would be a type statement, which allows users to declare a new custom type. It may contain type variables, but each of these must have a unique name:

 2.25 An example of type statement in Frege-IDE

Error detection like this is easily implemented by using a checking rule for the concept representing the whole type statement.

2.26 An example of a checking rule in Frege-IDE

A checking rule is basically a single method that is executed for a specific concept upon its any change. We tell MPS there is some semantic error by using “error” statement in the checking rule definition. “error” statement accepts a string (a message to display to the user of the language) and a node, which caused the problem. The MPS then underlines the corresponding node with red color in the editor to denote the error. *(Note that if a node’s reference is provided, then the original node is underlined, not the reference.)*

Typesystem rules, on the other hand, offer a declarative way to express rules which support type calculations. [FC] We can let MPS calculate types of expressions in the runtime and upon finding inconsistencies MPS will report error to the user on its own.

[FC – about w/s st:] MPS supports weak and strong sub-typing relations. A weak sub-type relation ta :< tb implies that expression of types ta and tb can be used interchangeably only in special cases. On the other hand a strong sub-type relation ta :<< tb implies ta :< tb and expressions of types ta and tb can be used interchangeably in all situations where expression of either ta or tb can be used. We can compare calculated types by using (both functions return boolean):

* isSubtype(type1 :< type2)
* isStrongSubtype(type1 :<< type2)

Inference rules are created to calculate a type of a node for a given concept. These can also be used to enforce a type, i.e. to perform a type check. An inference rule consists of a name, “applicable for” clause (where we specify a concept which we want to calculate a type for) and a “do“ part, where we define the rules for the current concept. The rules are written in the standard Java-like imperative language used also in the other MPS aspects (such as behavior), but it is extended with statements regarding the typesystem rules. An inference rule for a concept representing an integer literal would look like:

typeof(intLiteral) :==: <int>

On the left side, we have typeof(intLiteral). With this we are simply denoting a type of a node of a concept “IntLiteral”. With the operator :==: we are telling MPS that the type on the left must be the same as the type of the right operand. This is the rule we are specifying. *(Note that it is possible to perform a check only – this has to be specified in the inspector window in MPS for the current operator :==:.)*

Finally, the <int> part denotes a quotation. It is used when a language designer needs to create nodes of the concepts of a language (the language may be even the one he or she is now creating). A standard way of creating a node of a concept would be (an equivalent piece of code):

node<IntTypeNode> itn = new initialized node<IntTypeNode>();

typeof(intLiteral) :==: itn;

Anything displayed inside the quotations symbols (< … >) is what a node would look like as if an editor was used to edit the actual AST (in this case “int” is a textual representation of a node of the concept IntTypeNode). The quotations are most useful when dealing with larger ASTs, where a series of statements of “new initialized node<IntTypeNode>();” and child assignments would be difficult to read. Additionally, “antiquotations” may be used to include some dynamic content. These are denoted by symbols %( … )% inside quotations. For instance:

var firstTypeVar;

typeof(firstTypeVar) :==: <int>;

typeof(intBoolTuple) :==: <( %( firstTypeVar )%, bool )>;

Based on this, this is what an inference rule could look like for a not-expression node (standard “!” operator in Java, C, …):

2.27 An exemplar inference rule for “not-expression” concept

If the type of the original expression is not boolean, MPS will find out and underline the error for us: 2.28 An automatic type check for “not-expression”

Sometimes it is necessary to create custom types that are not present in jetbrains.mps.baseLanguage. For instance, consider a tuple in Frege language. A tuple is basically just a wrapper of 2 or more items of (not necessarily) different types. To create a tuple type, we have to create a new concept, which inherits from jetbrains.mps.baseLanguage.Type concept. We then specify its children. For simplicity, let us assume a tuple may contain 1 or more items rather than 2 or more. In that case, we declare that our “TupleType” concept has [1..n] children of type “jetbrains.mps.baseLanguage.Type”. To be able to use our new “TupleType” in typesystem rules, we have to include the language the concept is defined in, for the typesystem aspect.

 2.28 Specifying included languages for a typesystem aspect in an exemplar project

There are several other operators to use when specifying typesystem rules, but we will mention only the 2 following ones:

:<=: Tells MPS that the type on the left-hand side is a weak sub-type of the type on the right-hand side. [FC]

:~: Tells MPS that the type of operands on either side of the operator are weakly comparable. [FC]

The :~: is useful for the type definition of “equals” operator (“==”): we just specify that the operands on the left and right are weakly comparable and set the resulting type to bool:

typeof(equalsExpression.operandLeft) :~: typeof(equalsExpression.operandRight) ;

typeof(equalsExpression) :==: <bool>;

The custom defined types, however, are not implicitly comparable. Inequations are normally solved in MPS by applying subtyping rules. This means to compare types like “int” and “double”, we would first need to create a subtyping rule, which tells MPS that “int” is a subtype of “double”. For a more complicated type, like tuple, we would, however, need to create a replacement rule. These are used by MPS to solve inequations where subtyping rules may not be used. A replacement rule is always created for 2 concepts (which represent corresponding types) and has a defined body. For the given 2 incomparable types, MPS removes the inequation and simply executes the body in the corresponding replacement rule.

2.29 An example of a replacement rule for tuples in an exemplar project

To see how :<=: operator would be used, let us consider an inference rule for the concept representing binary + operation. Let us imagine that the operation is supported for all of the current types, but the types need to be comparable between each other, i.e. “double + int” is supported and the result is double, but “double + bool” is not. However, “bool + bool” is supported with the result “bool”, whereas “string + string” returns “string”, and so on. To express something like that in MPS, we can simply use the following piece of code:

infer typeof(plusExpression.operandLeft) :<=: typeof(plusExpression);

infer typeof(plusExpression.operandRight) :<=: typeof(plusExpression);

These two statements tell MPS that the result of the plus expression is the most “abstract” type of all of its two operands.

Now, to make things a little more complicated, we may assume our exemplar language on its highest level consists of series of statements each of which is a simple declaration of a variable, for instance:

i = (1 + 7.14)

j = (true == false)

k = !((1, 2.7) == (7, 2.7))

We can, however, use a declared variable as an expression, e.g.:

l = k + 1

With the current implementation, we would find out that MPS did not, in fact, underline the statement as an erroneous one *(“bool” + “int” is not defined)*. Why? The reason is that the type of “k” is not guaranteed to be evaluated before the type of “l”. It is different for expression (true == false), where the expression forms a binary tree and its type can be evaluated “bottom-up”. But the statements “l” and “k” are simply children of a root node and their ordering does not play any role here (we could swap statements “k“ and “l” and still expect MPS to evaluate the types correctly – just like in languages like Java and C#, where we can have a variable of a type that is declared later in the code). We can deal with the problem by using “when concrete” operator. “when concrete” has a form of a block of code and it delays its evaluation until the time the type of the requested argument has been calculated. The plus operator would then look like this:

when concrete (typeof(plusExpression.operandLeft) as olType) {

when concrete (typeof(plusExpression.operandRight) as orType) {

infer olType :<=: typeof(plusExpression);

infer orType :<=: typeof(plusExpression);

}

}

It is important to note that the code wrapped inside the “when concrete” operator is executed in a different thread, which means rest of the code of a currently defined typesystem rule will carry on with its execution independently.

* Chapter 3: Frege
  + *what is Frege*
  + *Frege vs. Haskell*
  + *Frege language - code examples, what it consists of (high-level) links to frege, github*
  + *Subset related to what we decided to support in “Frege-IDE”*

<https://github.com/Frege/frege> [FG - Frege github]

Frege is a Haskell for JVM. It is a purely functional language, has a strong static type system with global type inference and non-strict evaluation. [FG] The language compiles to Java and runs naturally on the JVM. This way it can be used inside any Java project. [FG]

The language was designed by Ingo Wechsung, who named it after the German mathematician, logician and philosopher Gottlob Frege. [WIKI]

<https://wiki.haskell.org/Foreign_Function_Interface#FFI_extension> <https://hackage.haskell.org/package/base-4.11.0.0/docs/Prelude.html>

<https://github.com/Frege/frege/wiki/Differences-between-Frege-and-Haskell> [1]

The syntax of Frege is very close to that of Haskell, with only small differences. In general, Frege could be considered a somewhat smaller subset of Haskell language. Certain features are missing, like the foreign function interface, which allows Haskell to interact with code written in another language. Instead, there are language constructs to make Java types and methods usable. All primitive types are just Java types, and so is String. [FG] The Frege-Prelude library (an equivalent of Haskell-Prelude library which defines many standard types like Maybe, functions, operators (||, &&), etc.) has many functions, type classes and types known from Haskell. However, Frege uses the Java APIs whenever possible, so certain points may feel different. [FG] For instance, implementation of type classes is incomplete and multi parameter type classes are not supported at all. [FG] Additionally, Frege does not support “newtype” declaration (an algebraic data type with exactly one constructor), “deriving” clause for data type declarations, and a few other keywords. A string value in Frege, unlike Haskell, is not a list of characters but an instance of the Java class java.lang.String. [FG] Also, Frege does not have any operator data constructor other than colon (“:”). This, however, allows to define even such custom operators that begin with the colon character (like :-: is a legal custom operator function in Frege). There are several other minor differences between Haskell and Frege, however, they mostly do not affect this work in any way, which is why we will not be describing them. To find out more information, please follow link [1].

**Frege features – later!**

Frege, based on Haskell language, has rather many syntactic (and semantic) constructs for this work to be able to include them all. We have therefore focused our attention only on the most important features worth examining, such as function declaration and definition, operators and (custom) datatypes. Our ideal IDE will have a user-friendly editor that should emulate normal text editing and writing code in the way that most Frege and Haskell developers are used to. This should be accompanied by a context help, sometimes referred to as ‘intellisense’, which would allow for referencing already defined functions, operators, variables, etc., in the corrects spots in the code. Last, but not least, we will strive for a type checker, which would be able to find small mistakes in the code, such as calling a function with illegal arguments, or evaluate type of an expression.

~~-~~

~~Frege, based on Haskell language, has rather many syntactic (and semantic) constructs for this work to be able to include them all.~~ We, therefore, focus only on the parts of the Frege that make it so popular functional language. ‘Syntactic sugars’ are for the most part omitted, as well as monads, which make Frege appear less functional and a more of an imperative programming language. To include more advanced features, like context help (references) and type system checking, we also had to keep the complexity of the work reasonably small (feasible) and thus concepts like classes and instances had to be omitted as well.

The following part includes the high level concepts of the supported Frege features.

(graphics / visuals.)

**module** fregeide.Skeleton

* A Frege program, or more specifically, a module, has to begin with a concept declaring its name. This allows for a module to be imported in other Frege modules while also making it clear to a developer (reading a program) what it does.

**module** fregeide.Skeleton(module fregeide.Submodule, ff, gg, hh)

* A module may specify which parts of the definition are visible to the outside (meaning: encapsulation). We may export functions, operators and declared datatypes. Additionally it is possible to re-export imported modules (see below).

**import** fregeide.Submodule

* We may import (one or more) modules inside the current one. This allows for usage of the imported functions, operators and datatypes.

**import** fregeide.Submodule as Sub (xx, yy)

* It is possible to specify which functions, operators and datatypes should be brought into the current namespace (we assume for simplicity there will be no naming conflicts), and which should be referenced by providing their full qualified names.

Infixr, function definition, annotation, type, data