# The STEX3 Manual \*

Michael Kohlhase, Dennis Müller FAU Erlangen-Nürnberg

http://kwarc.info/

2022-02-11

#### Abstract

STEX is a collection of LaTeX package that allow to markup documents semantically without leaving the document format, essentially turning LaTeX into a document format for mathematical knowledge management (MKM). STeX augments LaTeX with

- Semantic macros that denote and distinguish between mathematical concepts, operators, etc. independent of their notational presentation,
- A powerful module system that allows for authoring and importing individual fragments containing document text and/or semantic macros, independent of

   and without hard coding – directory paths relative to the current document,
- A mechanism for exporting ST<sub>E</sub>X documents to (modular) XHTML, preserving all the semantic information for semantically informed knowledge management services.

This is the user manual for the STEX package and associated software. It is primarily directed at end-users who want to use STEX to author semantically enriched documents. For the full documentation, see the STEX documentation

<sup>\*</sup>Version 3.0 (last revised 2022-02-11)

# Contents

1	What is STEX?	1
2	Quickstart	2
	2.1 Setup	2
	2.1.1 The STFX IDE	2
	2.1.2 Manual Setup	2
	2.2 A First ST <sub>E</sub> X Document	3
3	Using Semantic Macros	5
4	sTeX Archives	6
	4.1 The Local MathHub-Directory	6
	4.2 The Structure of STEX Archives	6
	4.3 MANIFEST.MF-Files	7
5	Creating New Modules and Symbols	8
	5.1 Advanced Structuring Mechanisms	8
	5.2 Primitive Symbols (The STEX Metatheory)	ç
6	STEX Statements (Definitions, Theorems, Examples,)	10
7	Additional Packages	11
1	7.1 Modular Document Structuring	11
		11
	7.3 Homework, Problems and Exams	11
8	Stuff	12
	8.1 Modules	12
	8.1.1 Semantic Macros and Notations	12
	Other Argument Types	14
	Precedences	15
	8.1.2 Archives and Imports	16
	Namespaces	16
	Paths in Import-Statements	17

# What is STEX?

Formal systems for mathematics (such as interactive theorem provers) have the potential to significantly increase both the accessibility of published knowledge, as well as the confidence in its veracity, by rendering the precise semantics of statements machine actionable. This allows for a plurality of added-value services, from semantic search up to verification and automated theorem proving. Unfortunately, their usefulness is hidden behind severe barriers to accessibility; primarily related to their surface languages reminiscent of programming languages and very unlike informal standards of presentation.

STEX minimizes this gap between informal and formal mathematics by integrating formal methods into established and widespread authoring workflows, primarily LATEX, via non-intrusive semantic annotations of arbitrary informal document fragments. That way formal knowledge management services become available for informal documents, accessible via an IDE for authors and via generated *active* documents for readers, while remaining fully compatible with existing authoring workflows and publishing systems.

Additionally, an extensible library of reusable document fragments is being developed, that serve as reference targets for global disambiguation, intermediaries for content exchange between systems and other services.

Every component of the system is designed modularly and extensibly, and thus lay the groundwork for a potential full integration of interactive theorem proving systems into established informal document authoring workflows.

The general STEX workflow combines functionalities provided by several pieces of software:

- $\bullet\,$  The STEX package to use semantic annotations in IATEX documents,
- RusTeX to convert tex sources to (semantically enriched) xhtml,
- The MMT software, that extracts semantic information from the thus generated xhtml and provides semantically informed added value services.

# Quickstart

#### 2.1 Setup

#### 2.1.1 The STEX IDE

TODO: VSCode Plugin

#### 2.1.2 Manual Setup

Foregoing on the STFX IDE, we will need several pieces of software; namely:

- The STEX-Package available here<sup>1</sup>. Note, that the CTAN repository for IATEX packages may contain outdated versions of the STEX package, so make sure, that your TEXMF system variable is configured such that the packages available in the linked repository are prioritized over potential default packages that come with your TEX distribution.
- The Mmt System available here<sup>2</sup>. We recommend following the setup routine documented here.
  - Following the setup routine (Step 3) will entail designating a MathHub-directory on your local file system, where the MMT system will look for STEX/MMT content archives.
- To make sure that STEX too knows where to find its archives, we need to set a global system variable MATHHUB, that points to your local MathHub-directory (see chapter 4).
- STEX Archives If we only care about IATEX and generating pdfs, we do not technically need MMT at all; however, we still need the MATHHUB system variable to be set. Furthermore, MMT can make downloading content archives we might want to use significantly easier, since it makes sure that all dependencies of (often highly interrelated) STEX archives are cloned as well.

Once set up, we can run mmt in a shell and download an archive along with all of its dependencies like this: lmh install <name-of-repository>, or a whole group of archives; for example, lmh install smglom will download all smglom archives.

 $<sup>^{1}\</sup>mathrm{EdNote}\colon$  For now, we require the latex3-branch

<sup>&</sup>lt;sup>2</sup>Ednote: For now, we require the sTeX-branch, requiring manually compiling the MMT sources

#### 2.2 A First STEX Document

Having set everything up, we can write a first STEX document. As an example, we will use the smglom/calculus and smglom/arithmetics archives, which should be present in the designated MathHub-folder.

The document we will consider is the following:

```
\documentclass{article}
\usepackage{stex}
\usepackage{stex}
\usepackage{xcolor}
\def\compemph#1{\textcolor{blue}{#1}}

\begin{document}
\usemodule[smglom/calculus]{series}
\usemodule[smglom/arithmetics]{realarith}

The \symref{series}{series} \sinfinitesum{n}{1}{
\text{realdivide}[frac]{1}{
\text{realpower}{2}{n}}
} \symref{converges}{converges} towards \$1\$.
\end{document}
```

Compiling this document with pdflatex should yield the output

```
The series \sum_{n=1}^{\infty} \frac{1}{2^n} converges towards 1.
```

Note that the  $\sum$  and  $\infty$ -symbols are highlighted in blue, and the words "series" and "converges" in bold. This signifies that these words and symbols reference STEX symbols formally declared somewhere; associating their presentation in the document with their (formal) definition - i.e. their semantics. The precise way in which they are highlighted (if at all) can of course be customized (see  $^3$ ).

\usemodule

The command \usemodule[some/archive] {modulename} finds some module in the appropriate archive – in the first case (\usemodule[smglom/calculus]{series}), STEX looks for the archive smglom/calculus in our local MathHub-directory (see chapter 4), and in its source-folder for a file series.tex. Since no such file exists, and by default the document is assumed to be in english, it picks the file series.en.tex, and indeed, in here we find a statement \begin{module}{series}.

STEX now reads this file and makes all semantic macros therein available to use, along with all its dependencies. This enables the usage of \infinitesum later on.

Analogously, \usemodule[smglom/arithmetics]{realarith} opens the file realarith.en.tex in the .../smglom/arithmetics/source-folder and makes its contents available, e.g. \realdivide and \realpower.

EdN:3

 $<sup>^3{</sup>m EdNote}$ : somewhere later

\symref \symname

The command \symref{symbolname}{text} marks the text in the second argument as representing the symbolname in the first argument – which is why the word "series" is set in boldface. In the pdf, this is all that happens. In the xhtml (which we will investigate shortly) however, we will note that the word "series" is now annotated with the full URI of the symbol denoting the mathematical concept of a series. In other words, the word is associated with an unambiguous semantics.

Notably, in both cases above (series and converges) the text that references the symbol and the name of the symbol are identical. Since this occurs quite often, the shorthand \symname{converges} would have worked as well, where \symname{foo-bar} behaves exactly like \symref{foo-bar}{foo bar} - i.e. the text is simply the name of the symbol with "-" replaced by a space.

\importmodule

If you investigated the contents of the imported modules (realarith and series) more closely, you'll note that none of them contain a symbol "converges". Yet, we can use \symref to refer to "converges". That is because the symbol converges is found in smglom/calculus/source/sequenceConvergence.en.tex, and series.en.tex contains the line \importmodule{sequenceConvergence}. The \importmodule-statement makes the module referenced available to all documents that include the current module. As such, a "current module" has to exist for \importmodule to work, which is why the command is only allowed within a module-environment.

TODO explain xhtml conversion, MMT compilation (requires an archive...?).

# Using Semantic Macros

TODO

# STEX Archives

#### 4.1 The Local MathHub-Directory

\usemodule, \importmodule, \inputref etc. allow for including content modularly without having to specify absolute paths, which would differ between users and machines. Instead, STEX uses archives that determine the global namespaces for symbols and statements and make it possible for STEX to find content referenced via such URIs.

All STEX archives need to exist in the local MathHub-directory. STEX knows where this folder is via one of three means:

- 1. If the STEX package is loaded with the option mathhub=/path/to/mathhub, then STEX will consider /path/to/mathhub as the local MathHub-directory.
- 2. If the mathhub package option is *not* set, but the macro \mathhub exists when the STEX-package is loaded, then this macro is assumed to point to the local MathHub-directory; i.e. \def\mathhub{/path/to/mathhub}\usepackage{stex} will set the MathHub-directory as path/to/mathhub.
- 3. Otherwise, STEX will attempt to retrieve the system variable MATHHUB, assuming it will point to the local MathHub-directory. Since this variant needs setting up only once and is machine-specific (rather than defined in tex code), it is compatible with collaborating and sharing tex content, and hence recommended.

#### 4.2 The Structure of STEX Archives

An STEX archive group/name needs to be stored in the directory /path/to/mathhub/group/name; e.g. assuming your local MathHub-directory is set as /user/foo/MathHub, then in order for the smglom/calculus-archive to be found by the STEX system, it needs to be in /user/foo/MathHub/smglom/calculus.

Each such archive needs two subdirectories:

- /source this is where all your tex files go.
- /META-INF a directory containing a single file MANIFEST.MF, the content of which we will consider shortly

An additional lib-directory is optional, and is where STEX will look for files included via \libinput.

Additionally a *group* of archives group/name may have an additional archive group/meta-inf. If this meta-inf-archive has a /lib-subdirectory, it too will be searched by \libinput from all tex files in any archive in the group/\*-group.

#### 4.3 MANIFEST.MF-Files

The MANIFEST.MF in the META-INF-directory consists of key-value-pairs, instructing STEX (and associated software) of various properties of an archive. For example, the MANIFEST.MF of the smglom/calculus-archive looks like this:

id: smglom/calculus

source-base: http://mathhub.info/smglom/calculus
narration-base: http://mathhub.info/smglom/calculus

 ${\tt dependencies: smglom/arithmetics,smglom/sets,smglom/topology,}$ 

smglom/mv,smglom/linear-algebra,smglom/algebra

responsible: Michael.Kohlhase@FAU.de

title: Elementary Calculus

teaser: Terminology for the mathematical study of change.

description: desc.html

Many of these are in fact ignored by STFX, but some are important:

id: The name of the archive, including its group (e.g. smglom/calculus),

source-base or

ns: The namespace from which all symbol and module URIs in this repository are formed, see (TODO),

narration-base: The namespace from which all document URIs in this repository are formed, see (TODO),

url: The URL that is formed as a basis for external references, see (TODO),

dependencies: All archives that this archive depends on. SIEX ignores this field, but MMT can pick up on them to resolve dependencies, e.g. for lmh install.

# Creating New Modules and Symbols

#### TODO

#### 5.1 Advanced Structuring Mechanisms

Given modules:

# | Learn | Lear

We can form a module for *rings* by "cloning" an instance of group (for addition) and monoid (for multiplication), respectively, and "glueing them together" to ensure they share the same universe:

#### Example 2

```
\begin{module}{ring}
\begin{copymodule}{group}{addition}
\renamedec[name=universe]{universe}{runiverse}
\renamedec[name=plus]{operation}{rplus}
\renamedec[name=zero]{unit}{rzero}
\renamedec[name=zero]{unit}{rzero}
\renamedec[name=uminus]{inverse}{ruminus}
\end{copymodule}
\notation[plus,op=+,prec=60]{rplus}{#1 \comp+ #2}
\notation[zero]{rzero}{\comp0}
\notation[uminus,op=-]{ruminus}{\comp- #1}
\begin{copymodule}{monoid}{multiplication}
\assign{universe}{\comp0}{renamedec[name=times]{operation}{rtimes}
\renamedec[name=one]{unit}{rone}
\end{copymodule}
\notation[cdot,op=\cdot,prec=50]{rtimes}{#1 \comp\cdot #2}
\notation[cdot,op=\{comp1}
\rest: $\rimes a{\rplus c{\rimes de}}$$
\end{module}
```

```
Module 5.1.4 [\text{ring}]
Test: a \cdot (c + d \cdot e)
```

#### TODO: explain donotclone

#### Example 3

```
\begin{module}{int}
\symdef{Integers}{\comp{\mathbb Z}}
\symdef{args=2,op=+|{plus}{#1 \comp+ #2}
\symdef{args=2,op=-|{minus}{\comp-#1}

\symdef{args=1,op=-|{minus}{\comp-#1}}

\begin{interpret module}{group}{intisgroup}
\assign{universe}{\Integers}
\assign{operation}{\plus!}
\assign{operation}{\plus!}
\assign{unit}{\zero}
\assign{inverse}{\uminus!}
\end{interpret module}
\end{module}
\end{module}
\end{module}
\end{module}
\end{\text{module}}
\text{module}
\text{module}}
\end{\text{module}}
\text{module}
\text{module}
```

```
Module 5.1.5[int]
```

## 5.2 Primitive Symbols (The STEX Metatheory)

STEX Statements (Definitions, Theorems, Examples, ...)

# **Additional Packages**

- 7.1 Modular Document Structuring
- 7.2 Slides and Course Notes
- 7.3 Homework, Problems and Exams

# Stuff

#### 8.1 Modules

\sTeX \stex

Both print this  $ST_EX$  logo.

#### 8.1.1 Semantic Macros and Notations

Semantic macros invoke a formally declared symbol.

To declare a symbol (in a module), we use \symdecl, which takes as argument the name of the corresponding semantic macro, e.g. \symdecl{foo} introduces the macro \foo. Additionally, \symdecl takes several options, the most important one being its arity. foo as declared above yields a *constant* symbol. To introduce an *operator* which takes arguments, we have to specify which arguments it takes.

For example, to introduce binary multiplication, we can do \symdecl[args=2]{mult}. We can then supply the semantic macro with arbitrarily many notations, such as \notation{mult}{#1 #2}.

```
Example 4

| \symdec! [args=2]{mult} \
    \notation \{ mult \{ #1 #2 \}} \\
    \mult \{ a \} \{ b \} \$

| ab
```

Since usually, a freshly introduced symbol also comes with a notation from the start, the \symdef command combines \symdecl and \notation. So instead of the above, we could have also written

```
\symdef[args=2]{mult}{#1 #2}
```

Adding more notations like  $\notation[cdot]{mult}{#1 } or \\ notation[times]{mult}{#1 } comp{\times} #2} allows us to write <math>\notation[times]{a}{b}$  and  $\notation[times]{a}{b}$ :

#### Example 5

.

EdN:4

Not using an explicit option with a semantic macro yields the first declared notation, unless changed  $^4$ .

Outside of math mode, or by using the starred variant \foo\*, allows to provide a custom notation, where notational (or textual) components can be given explicitly in square brackets.

#### Example 6

.

In custom mode, prefixing an argument with a star will not print that argument, but still export it to OMDoc:

#### Example 7

The syntax  $*[\langle int \rangle]$  allows switching the order of arguments. For example, given a 2-ary semantic macro \forevery with exemplary notation \forall #1. #2, we can write

#### Example 8

.

When using \*[n], after reading the provided (nth) argument, the "argument counter" automatically continues where we left off, so the \*[1] in the above example can be omitted.

<sup>&</sup>lt;sup>4</sup>EdNote: TODO

For a macro with arity > 0, we can refer to the operator *itself* semantically by suffixing the semantic macro with an exclamation point! in either text or math mode. For that reason \notation (and thus \symdef) take an additional optional argument op=, which allows to assign a notation for the operator itself. e.g.

#### Example 9

```
The operator + adds two elements, as in a + b
```

\* is composable with! for custom notations, as in:

#### Example 10

```
\mult![\comp{Multiplication}] (denoted by $\mult*![\comp\cdot|$) is defined by...
Multiplication (denoted by ·) is defined by..
```

The macro \comp as used everywhere above is responsible for highlighting, linking, and tooltips, and should be wrapped around the notation (or text) components that should be treated accordingly. While it is attractive to just wrap a whole notation, this would also wrap around e.g. the arguments themselves, so instead, the user is tasked with marking the notation components themself.

The precise behaviour of \comp is governed by the macro \@comp, which takes two arguments: The tex code of the text (unexpanded) to highlight, and the URI of the current symbol. \@comp can be safely redefined to customize the behaviour.

The starred variant \symdecl\*{foo} does not introduce a semantic macro, but still declares a corresponding symbol. foo (like any other symbol, for that matter) can then be accessed via \STEXsymbol{foo} or (if foo was declared in a module Foo) via \STEXModule{Foo}?{foo}.

both \STEXsymbol and \STEXModule take any arbitrary ending segment of a full URI to determine which symbol or module is meant. e.g. \STEXsymbol{Foo?foo} is also valid, as are e.g. \STEXModule{path?Foo}?{foo} or \STEXsymbol{path?Foo?foo}

There's also a convient shortcut \symref{?foo}{some text} for \STEXsymbol{?foo}! [some text]

#### Other Argument Types

So far, we have stated the arity of a semantic macro directly. This works if we only have "normal" (or more precisely: i-type) arguments. To make use of other argument types, instead of providing the arity numerically, we can provide it as a sequence of characters representing the argument types – e.g. instead of writing args=2, we can equivalently write args=ii, indicating that the macro takes two i-type arguments.

Besides i-type arguments, STFX has two other types, which we will discuss now. The first are binding (b-type) arguments, representing variables that are bound by

the operator. This is the case for example in the above \forevery-macro: The first

argument is not actually an argument that the forevery "function" is "applied" to; rather, the first argument is a new variable (e.g. x) that is bound in the subsequent argument. More accurately, the macro should therefore have been implemented thusly:

```
\symdef[args=bi]{forevery}{\forall #1.\; #2}
```

b-type arguments are indistinguishable from i-type arguments within STFX, but are treated very differently in OMDoc and by MMT. More interesting within STEX are a-type arguments, which represent (associative) arguments of flexible arity, which are provided as comma-separated lists. This allows e.g. better representing the \mult-macro above:

#### Example 11

```
a \cdot b \cdot c \cdot d^e \cdot f
```

As the example above shows, notations get a little more complicated for associative arguments. For every a-type argument, the \notation-macro takes an additional argument that declares how individual entries in an a-type argument list are aggregated. The first notation argument then describes how the aggregated expression is combined into the full representation.

For a more interesting example, consider a flexary operator for ordered sequences in ordered set, that taking arguments {a,b,c} and \mathbb{R} prints  $a \leq b \leq c \in \mathbb{R}$ . This operator takes two arguments (an a-type argument and an i-type argument), aggregates the individuals of the associative argument using \leq, and combines the result with \in and the second argument thusly:

#### Example 12

```
a \le b \le c \in \mathbb{R}
```

Finally, B-type arguments combine the functionalities of a and b, i.e. they represent flexary binding operator arguments.

#### Precedences

Every notation has an (upwards) operator precedence and for each argument a (downwards) argument precedence used for automated bracketing. For example, a notation for a binary operator \foo could be declared like this:

```
\noindent [prec=200;500x600]{foo}{#1 \comp{+} #2}
```



 $<sup>^5{\</sup>rm EdNote}$  what about e.g. \int \_x\int \_y\int \_z f dx dy dz?

<sup>&</sup>lt;sup>6</sup>EDNOTE: "decompose" a-type arguments into fixed-arity operators?

assigning an operator precedence of 200, an argument precedence of 500 for the first argument, and an argument precedence of 600 for the second argument.

SIEX insert brackets thusly: Upon encountering a semantic macro (such as \foo), its operator precedence (e.g. 200) is compared to the current downwards precedence (initially \neginfprec). If the operator precedence is *larger* than the current downwards precedence, parentheses are inserted around the semantic macro.

Notations for symbols of arity 0 have a default precedence of \infprec, i.e. by default, parentheses are never inserted around constants. Notations for symbols with arity > 0 have a default operator precedence of 0. If no argument precedences are explicitly provided, then by default they are equal to the operator precedence.

Consequently, if some operator A should bind stronger than some operator B, then As operator precedence should be smaller than Bs argument precedences.

For example:

#### Example 13

```
\notation[prec=100]{plus}{#1 \comp{+} #2}
\notation[prec=50]{times}{#1 \comp{\cdot} #2}
$\plus{a}{\times{b}{c}}$ and $\times{a}{\plus{b}{c}}$
```

8.1.2 Archives and Imports

#### Namespaces

Ideally, STEX would use arbitrary URIs for modules, with no forced relationships between the *logical* namespace of a module and the *physical* location of the file declaring the module – like MMT does things.

Unfortunately, TEX only provides very restricted access to the file system, so we are forced to generate namespaces systematically in such a way that they reflect the physical location of the associated files, so that STEX can resolve them accordingly. Largely, users need not concern themselves with namespaces at all, but for completenesses sake, we describe how they are constructed:

- If \begin{module}{Foo} occurs in a file /path/to/file/Foo[.\(\lang\)].tex which does not belong to an archive, the namespace is file://path/to/file.
- If the same statement occurs in a file /path/to/file/bar[.\(\lang\)].tex, the namespace is file://path/to/file/bar.

In other words: outside of archives, the namespace corresponds to the file URI with the filename dropped iff it is equal to the module name, and ignoring the (optional) language suffix<sup>1</sup>.

If the current file is in an archive, the procedure is the same except that the initial segment of the file path up to the archive's source-folder is replaced by the archive's namespace URI.

<sup>&</sup>lt;sup>1</sup>which is internally attached to the module name instead, but a user need not worry about that.

#### Paths in Import-Statements

Conversely, here is how namespaces/URIs and file paths are computed in import statements, examplary \importmodule:

- \importmodule{Foo} outside of an archive refers to module Foo in the current namespace. Consequently, Foo must have been declared earlier in the same document or, if not, in a file Foo[.\lang\rangle].tex in the same directory.
- The same statement within an archive refers to either the module Foo declared earlier in the same document, or otherwise to the module Foo in the archive's top-level namespace. In the latter case, is has to be declared in a file Foo[. $\langle lang \rangle$ ].tex directly in the archive's source-folder.
- Similarly, in \importmodule{some/path?Foo} the path some/path refers to either the sub-directory and relative namespace path of the current directory and namespace outside of an archive, or relative to the current archive's top-level namespace and source-folder, respectively.
  - The module Foo must either be declared in the file  $\langle top\text{-}directory \rangle$ /some/path/Foo[. $\langle lang \rangle$ ].tex, or in  $\langle top\text{-}directory \rangle$ /some/path[. $\langle lang \rangle$ ].tex (which are checked in that order).
- Similarly, \importmodule[Some/Archive]{some/path?Foo} is resolved like the previous cases, but relative to the archive Some/Archive in the mathhub-directory.
- Finally, \importmodule{full://uri?Foo} naturally refers to the module Foo in the namespace full://uri. Since the file this module is declared in can not be determined directly from the URI, the module must be in memory already, e.g. by being referenced earlier in the same document.
  - Since this is less compatible with a modular development, using full URIs directly is discouraged.