ECP-2008-EDU-428046

Math-Bridge

Content Release

Deliverable number	*D-2.1*
Dissemination level	*Confidential*
Delivery date	*31 July 2010*
Status	*Draft*
Authors	*Eric Andrès (Saarland University)* *Michael Dietrich (Saarland University)*

e*Content*plus

This project is funded under the *e*Content*plus* programme[1] ,
a multiannual Community programme to make digital content in Europe more accessible, usable and exploitable.

[1]OJ L 79, 24.3.2005, p. 1.

Project ref.no.	ECP-2008-EDU-428046
Project title	Math-Bridge European Remedial Content for Mathematics

Deliverable dissemination level	Confidential
Contractual date of delivery	July 31^{th}, 2010
Actual date of delivery	July 31^{th}, 2010
Deliverable number	D-2.1
Deliverable title	Content Release
Type	Deliverable
Status & version	Draft
Number of pages	27
WP contributing to the deliverable	WP2
WP/Task responsible	Eric Andrès/Michael Dietrich
Authors	*Eric Andrès (Saarland University)* *Michael Dietrich (Saarland University)*
EC Project Officer	Marcel Watelet
Keywords	content, quality management, knowledge representation

Contents

1 Source Content **6**

1.1 German Research Center For Artificial Intelligence 6
1.2 University of Kassel/University of Paderborn 6
1.3 Open Universiteit Netherlands 7
1.4 Saarland University 8
1.5 University of Vienna 8
1.6 Tampere University of Technology 8

2 Transformation Steps **11**

2.1 Slicing 11
2.2 Structure Annotation 11
2.3 Metadata 11
2.4 Add Semantics 12
2.5 tex2omdoc, a batch script approach 12
2.6 jEditOQMath's smart-paste approach 13

3 Quality Management **14**

3.1 The Gap Detection Tool 14
3.2 Current State of content review 16

4 Current State **17**

4.1 German Research Center For Artificial Intelligence 17
4.2 University of Kassel/University of Paderborn 17
4.3 Open Universiteit Netherlands 18
4.4 Saarland University 18
4.5 University of Vienna 19
4.6 Tampere University of Technology 19

5 Summary **20**

A Content Authors **21**

A.1 German Research Center for Artificial Intelligence 21

A.2 University of Kassel/University of Paderborn 21

A.3 Open Universiteit Netherlands 22

A.4 Saarland University 23

A.5 University of Vienna 23

A.6 Tampere University of Technology 23

B Review Criteria **25**

B.1 Minimal Requirements 25

B.2 Technical Review Criteria 25

B.3 Content Review Criteria 26

B.4 Pedagogical Review Criteria 27

Introduction

In order to wake students or teachers interest, intelligent tutoring systems (ITSs) need mainly one thing besides advanced technical features like adaptation to student knowledge or interactive exercises: high quality content.

The creation of such content is a time consuming and therefore costly process. In this report we describe all the steps that were performed to create the content that is used in the Math-Bridge project, beginning with a description of the content sources we started with. Then we name and explain all transformation steps that were performed to convert the sources to the target format as well as developments that were made to support these transformations. In the following section, we describe the workflow for quality management we applied to ensure and achieve the desired high quality of the content. We conclude with a summary

1 Source Content

Content plays an important role in e-learning systems. It is the main thing the user is interacting with when using the system. If the user cannot find content on a needed topic or the content quality is too poor the system will not be used unless someone(teacher/parents) forces her. On the other hand an e-learning system with high quality content will be used happily and the word on the quality will spread. To be able to cover a lot of topics with high quality content, six partners of the Math-Bridge consortium provided content that has already widely been used for teaching, especially in the context of bridging courses.

In the next sections we describe the content provided by each of these partners in more detail.

1.1 German Research Center For Artificial Intelligence

The German Research Center for Artificial Intelligence (DFKI) contributed a course on fraction arithmetics to the Math-Bridge content pool. This course has originally been created by Edgar Kessler, a teacher who worked for almost two years with the ActiveMath group. It has been extended and improved over time. The course aims at middle school students, but can also be used by first semester students to refresh their knowledge on the topic. Many teachers, spread all over Germany, used this course to support their teaching of fractions. The content was delivered in OMDoc, the format needed by ActiveMath, in German language. Although the content was already in the appropriate format, there were still some adjustments needed which will be described later in section 2. It contains about 700 learning objects.

1.2 University of Kassel/University of Paderborn

The partners from University of Kassel/Paderborn(KS-PB) provided an advancement of their VEMA [2] materials. The development of these materials started March 2003 and undergoes steady evaluations and improvements according to student feedback. There are many authors involved of whom *Prof. Dr. Rolf Biehler*, *Prof. Dr. Reinhard Hochmuth*, *Pascal Fischer* and *Thomas Wassong* are members of the Math-Bridge consortium. A full list of can be seen in A.2.

Their material is designed for pre- and first-grade-students for engineering, science, computer science and mathematics as well as for pre-service-teachers and covers the following topics:

[2] see website `http://www.mathematik.uni-kassel.de/~vorkurs/`

- Basic arithmetic
- (In-)equalities
- Sets
- Logic
- Powers, quadratic equalities
- Linear, quadratic and exponential functions and polynomials
- Logarithm
- Sequences and series
- Calculus and integral calculus

The material was provided as LaTeX documents, which resulted in roundabout 700 printed A4 pages. To support fast transformation of this materials the tex2omdoc (see 2.5) tool was developed. The content was provided in two languages: English and German.

1.3 Open Universiteit Netherlands

The Open Universiteit Netherlands (OUNL) contributed five collections from their teaching materials. All these collections are suitable for first or second year students in engineering. Again there is a huge set of authors involved of which *Evert van de Vrie* and *Josje Lodder* are members of the Math-Bridge consortium. A full list of authors and contributors can be seen in Appendix A.3. The material covers the following topics:

- Complex numbers and complex plane
- Polar coordinates and complex powers of e
- Second order linear differential equations
- Practical computer algebra
- Differential equations and direction fields
- First order differential equations
- Sets of differential equations
- Calculus

- Gaussian elimination
- Vectors and matrices
- Matrix algebra and determinants
- Linear optimisation

The collections were provided as pdf documents, which would make a pile of 300 printed A4 pages in Dutch.

1.4 Saarland University

Saarland University provided its calculus course which originates from "LeActiveMath" [3], an EU funded project. It was originally authored by Marianne Moormann and Christian Groß. See an exhaustive list of authors in Appendix A.4. In the course of ActiveMath-EU[4], another EU funded follow-up project, the course has been fully translated from German and English to Czech, Spanish and Hungarian. Partial translations have been done for Dutch, French and Mandarin Chinese. As the course was authored for the ActiveMath system it was already in the needed format, but similar to the fractions course provided by DFKI, some modifications were needed to make the course suitable for the new functionalities planned for Math-Bridge. It contains of around 1600 learning objects.

1.5 University of Vienna

Our partner from Austria, the University of Vienna provided the contents of `http://www.mathe-online.at`, which is a huge collection of mathematical content covering many topics for learners from middle-school to university. The development originates back to march 1998 with *Franz Embacher* and *Petra Oberhuemer* as main authors. Many persons contributed to this material, which can be seen in Appendix A.5.

The content was provided as HTML pages in German language.

1.6 Tampere University of Technology

The Finnish partners from Tampere University of Technology (TUT) contributed two collections to the Math-Bridge content pool. One of these collections is used in Finland as remedial material for Finnish high-school mathematics. *Seppo Pohjolainen* is among the

[3] See LeActiveMath project website: `http://www.leactivemath.org/`
[4] See ActiveMath-EU project website: `http://www.activemath.org/eu/`

authors of this collection. Appendix A.6 shows a list of authors. The collection covers the following topics:

- Set theory
- Logic
- Real numbers
- Expressions
- Functions
- Linear equations and inequalities
- Trigonometry
- Complex numbers
- Vectors
- Equations and inequalities
- Analytic geometry
- Logarithmic and exponential functions
- Differential calculus
- Integral calculus
- Sequences and series

This collection was provided as HTML pages in Finnish language.

The second collection aims at first year students in Finland and covers the following topics:

- Basics of set theory, logic and proving
- General function theory and elementary functions
- Complex numbers
- Vectors
- Linear system of equations

- Matrices
- Eigenvalues and eigenvectors
- Orthogonality
- Limit of a function and continuity
- Derivative
- Integral calculus
- Curves
- Differential calculus of multi-variable functions

This materials were provided as pdf documents in Finnish language

After this quick overview of the content sources we explain which tasks had to be performed in order to make the material usable for the Math-Bridge project.

2 Transformation Steps

There were three different formats (HTML, PDF and LaTeX) which had to be transformed to OMDoc in order to be used by ActiveMath. In this section we describe the four steps that were performed:

2.1 Slicing

The source material provided by the partners consists of coarse grained learning units, which makes them hard to reuse. The Math-Bridge service provides adaptive content assembled from all available materials, so it needs access to fine-grained learning objects that are well reusable. In order to achieve this, the big learning units needed to be sliced into smaller units, the so-called learning objects (LOs). If the source format provides enough "metadata" to tell that a certain part of text represents a learning object, this slicing process can be done semi-automatically. An environment in LaTeX is an example of such metadata. As the content of Kassel/Paderborn was provided in LaTeX we could partially reuse the implicit slicing provided in the form of custom environments. For all other content except the material provided by USAAR and DFKI the slicing had to be done completely manually. OUNL provided slicing of their material; the other collections were sliced by the content encoders (the persons doing the transformation of the content) and checked during the quality control by the providing partner.

2.2 Structure Annotation

Learning objects can be of different types in ActiveMath. For example one learning object can be a definition, another one an exercise or proof. After slicing, we had to assign a type to each of the sliced learning objects. This step had to be performed for all collections except the ones provided by USAAR and DFKI as this was already done.

2.3 Metadata

In order for the Math-Bridge service to automatically assemble courses tailored to the users needs, we need to provide certain information on learning objects to allow the tutorial component to discover appropriate ones. For this purpose three types of metadata are used: descriptive, pedagogic and semantic metadata.

- Descriptive metadata provides the data necessary to properly catalogue and discover a knowledge item.

- Pedagogic metadata describe pedagogic properties of knowledge items
- Semantic Metadata are used to structure the subject of study and learning material.

The learning objects of all collections had to be annotated with metadata. This also holds for the content collections provided by DFKI and USAAR as their metadata schemes are different from the one used in Math-Bridge.

2.4 Add Semantics

Mathematical texts contain a lot of formulæ. In the source material, these formulæ are mostly represented as text or even worse, as images. ActiveMath uses a semantic representation for formulæ allowing features like formula search or copy-and-paste of formulas. So as a final step of content transformation all these formulæ needed to be translated to the semantic representation used in ActiveMath. This transformation step had to be done for all new collections and partially for the already existing collection provided by DFKI and USAAR as it turned out, that some formulæ weren't encoded at all or not done in a proper way.

For this semantic enrichment, heuristics based procedures are needed, they are of two sorts: batch script and smart paste.

2.5 tex2omdoc, a batch script approach

The batch script approach relies on parsing the source format, it has been applied for the LaTeX-encoded sources of Kassel and Paderborn partners. The tool uses the LaTeXML processor,[5] a converter which produces an XML representation of the LaTeX source. In a typical application of LaTeXML, the intermediate XML representation is transformed to a target format ,e.g., HTML+MathML using a post-processor provided by the author of LaTeXML.

tex2omdoc uses a similar approach by applying an XSLT transformation pipeline on the XML produced by LaTeXML. The intermediate XML format had to be extended to cope with the expressive power of the target language. For instance, we needed to add language elements representing the different types of learning objects. The LaTeXML processor itself had also to be extended and adapted to recognize the custom environments used by Kassel/Paderborn and map them to the extended format. The XSLT transformation pipeline makes heavy use of heuristics to determine formula semantics, because even well structured LaTeX code is still presentation-oriented. A fully automatic conversion was not possible, but a slow maturation process of the converter yielded an OQMath output with sufficient semantics after which a manual cleanup was performed.

[5] LaTeXML is an open-source processor written in Perl and XSLT. See `http://dlmf.nist.gov/LaTeXML/`.

2.6 jEditOQMath's smart-paste approach

The smart paste approach uses less structured sources and relies on more manual work: based on sources in such formats as PDF, Word, or HTML, a copy operation is requested obtaining at least the plain text which can be pasted within the authoring environment, jEditOQMath.[6] Each of the ingredients can then be obtained by a dedicated copy (e.g. *the Copy MathML* function of Firefox, the copy of the image which can copy the TeX source as plain text). The smart-paste function of jEditOQMath then combines a set of heuristics to convert to a semantic form the formula sources; it uses several converters which it all attempts giving results which the user can choose from.[7] This approach works for formulæ and for pictures and has the biggest advantage to allow adjustments on the fly before the conversion is applied.

Each of these conversion processes deliver OQMath files which the encoder will critically check within his authoring environment. His checks generally give rise to further adjustments to the source and/or to the encoding process.

This section described the steps content material needs to undergo to be really useful in ActiveMath. The next section describes an approach on how to ensure the quality of the transformed content.

[6] jEditOQMath is an open-source combination of supported XML source editing and communication tools to ActiveMath, see `http://www.activemath.org/projects/jEditOQMath`.

[7] The smart paste function is explained in the MathUI paper `http://www.activemath.org/workshops/MathUI/09/`.

3 Quality Management

In the preceding section we described the steps that are needed for transforming content from some source format to OMDoc, the format used by ActiveMath. In each of these steps errors can be produced. Unfortunately, only very few of them can be detected automatically, which makes it necessary to manually check the content.

We describe the approach used to ensure the quality of the content used in Math-Bridge. As the approach is already in use, a small description of the current state of quality management is provided here. We first describe a tool that has been developed for finding errors:

3.1 The Gap Detection Tool

The Gap Detection Tool has been developed during the first year of the Math-Bridge project. Its purpose is to find the following three different kinds of errors in content collections.

- Structural Gaps
- Language Gaps
- Didactic Gaps

Structural gaps show wrong or missing relations between learning objects. A relation is considered to be wrong when it points to a learning object that is not existing or to itself. To ensure that all components of ActiveMath can work properly with the content these relations need to be fixed.

Language gaps report learning objects where the translation has only partially been done or is not done at all. Of course we are interested in having all learning objects translated to be able to reach a broader audience.

Didactic gaps occur when the course generation is used. The course generation uses different pedagogical scenarios to assemble books tailored to the users needs. These scenarios request a sequence of learning object which fit some metadata. A problem occurs when a requested learning object is not available. In such a situation, the course generation will relax the requirements (if possible) and ask for another learning object. If none can be found it will again try to relax its requirements and finally leave out this learning object if it cannot find a suitable one.

Every time the course generation needs to relax its requirements (or in the worst case skips a request completely), a didactic gap is found. The Gap Detection Tool has been connected with the tutorial component and keeps track of all found gaps. To support authors it is capable of providing a skeleton for the missing learning object, giving an author the opportunity to create content suiting the needs of the tutorial component. With the didactical gap detection, we ensure high quality automatically assembled books.

All found gaps are stored in a database inside the server installation, where it is also possible to manually add some gaps found by a user.

The Gap Detection has been integrated in our review workflow which consists of four parts and involves four groups of people. For every content collection used in Math-Bridge we assign an review organizer, who is solely responsible for the review of that collection. Then we divide the reviewers in three different groups. Technical reviewers have the task to check the technical correctness of the produced content. Usually these are the same persons that transform the content to OMDoc. The organizer has to ensure they do not review their own encoding work. Content reviewers check the correctness of the content, e.g., whether there are typographic errors in the content or whether formulæ are mathematically correct. This group consists mostly of content providing partners, because the contents are often in languages the encoders cannot understand.

The last group are the pedagogical reviewers who take care that books created using the tutorial component are didactically correct and useful. This group consists of pedagogists from the project partners or persons outside the consortium.

Each of the review groups checks a set of criteria listed in section **??** according to the following workflow:

1. The review organizer is informed by the person in charge of the content encoders that his collection is ready for review. He is responsible to check that the minimal requirements are really met. If this is not the case the review organizer reports the problems by creating issues in Jira **??**, the bug-tracking system used in Math-Bridge.
2. If the minimal requirements are met (or all reported issues are fixed) the review organizer distributes review tasks using Jira to the technical review team.
3. The technical review team reports its findings to the review organizer by adding pages to the Math-Bridge wiki. They report the completion of their review tasks by closing the Jira issue assigned to them.
4. From the reported findings the review organizer again creates issues in Jira, which are assigned to the content encoders. As soon as all issues are fixed the review organizer has the technical team review the collection again, to make sure that no problem was overseen.
5. Concurrently he creates review issues for the content review team.

6. The content reviewers also report problems using wiki pages. Similar to the technical review team the content review team reports completion of work by closing the issue.
7. From the reported problems appropriate issues are created for the content encoding team.
8. As soon as the content encoding team marked all problems as resolved the review organizer will again order the content review team to re-check the collection and thus make sure no error was overseen.
9. Once this has been done for all collections used in Math-Bridge the pedagogical review team starts its work by using the course generation to create automatically assembled books and check if they are didactically correct and useful for the users. Should there be a problem, a joint effort of pedagogists, course generation scenario developers and content encoders will be needed to fix it.

3.2 Review Criteria

In this section, we list the criteria used in review workflow.

3.2.1 Minimal Requirements

- parse error free
- duplicate id error free
- (mostly) dtd error free
- contains no unresolved xrefs (for and domain-prerequisite relations, links)
- language set correctly
- are images used to present formulæ
- (all) formulæ qmath/openmath encoded
- Metadata like author/contributor should be present
- before assigning review task to content providers
- a recbook hiding symbols should be available
- Metadata for course generation to work (learningcontext)
- for and domain_prerequisite should be given
- Gap detector gives only acknowledged gaps

3.2.2 Technical Review Criteria

Primary

- does LO contain errors (i.e. transformation errors)
- is LO layout correct according to source
- are tooltips working (display—correct semantic—language)
- are multimedia objects presented
- are images used to present formulæ
- are formulæ rendered correct
- is book page printable?
- is math-search and other search expectedly working?

Exercise

- the exercises runs without errors with input-editor and with linear input

Secondary

- are formulæ mathematically correct
- are country specific notations used
- are formulæ that one expects to drag-and-drop actually droppable?

Exercise

- correct solutions can be input (need to read the source for it)
- is a diagnose sent to user-model? (see mastery bullet)

3.2.3 Content Review Criteria

Primary

- is wording correct and consistent
- are formulæ mathematically correct
- are country specific notations used
- is slicing correct
- are links correct
- is order of LOs correct
- is structure of book(s) correct
- are there LOs which don't fit
- are LOs missing?
- check correctness of relations between individual LOs
- are LOs findable in search

Secondary

- does LO contain errors (ie transformation errors)
- is multimedia presented
- are formulæ rendered correct

Exercises

- exercises can be run with the correct input as expected mathematically
- exercises provide feedback for "normally expected errors"

3.2.4 Pedagogical Review Criteria

Primary

- Is course generation generating sensible books?
- Does search produce all and only all relevant results?
- are successful exercise runs reporting a positive achievement and score?
- are unsuccessful exercise runs reporting a bad score?

4 Current State

In this section we provide a description of the current state of the released content. This version of the content can be used online at `http://service.math-bridge.org`. In the following tables we provide information on how far the transformation processes have come, name the number of resulting learning objects(LOs) if applicable and give information on the languages the content is available in.

4.1 German Research Center For Artificial Intelligence

Collection Name: Fractions	Encoding Partner	USAAR
Content Transformation	DONE	
Content Slicing	DONE	
Number of LOs	642	
Available Languages	German	642 LOs
Metadata Enrichment	IN PROGRESS	
Formula Transformation	DONE	
Quality Control	Technical review ongoing	

4.2 University of Kassel/University of Paderborn

Collection Name: Kassel Content	Encoding Partner	USAAR
Content Transformation	IN PROGRESS	
Content Slicing	IN PROGRESS	
Number of LOs	183	
Available Languages	German	183 LOs
Metadata Enrichment	IN PROGRESS	
Formula Transformation	IN PROGRESS	
Quality Control	Technical, content review ongoing	

4.3 Open Universiteit Netherlands

Collection Name: ounlContent	Encoding Partner	USAAR
Content Transformation	DONE	
Content Slicing	DONE	
Number of LOs	822	
Available Languages	Dutch	822 LOs
Metadata Enrichment	IN PROGRESS	
Formula Transformation	DONE	
Quality Control	Technical review ongoing	

4.4 Saarland University

Collection Name: LeAM-calculus	Encoding Partner	USAAR
Content Transformation	DONE	
Content Slicing	DONE	
Number of LOs	1618	
Available Languages	English	1618 LOs
	German	1618 LOs
	Hungarian	1618 LOs
	Spanish	1618 LOs
	French	729 LOs
	Dutch	71 LOs
Metadata Enrichment	DONE	
Formula Transformation	DONE	
Quality Control	Technical review done, content review pending	

4.5 University of Vienna

Collection Name:	Encoding Partner	ELTE
Content Transformation		
Content Slicing		
Number of LOs		
Available Languages	German English	
Metadata Enrichment		
Formula Transformation		
Quality Control		

4.6 Tampere University of Technology

Collection Name: tutContent	Encoding Partner	USAAR
Content Transformation	DONE	
Content Slicing	DONE	
Number of LOs	1855	
Available Languages	Finnish	1855 LOs
Metadata Enrichment	IN PROGRESS	
Formula Transformation	DONE	
Quality Control	Technical, content review ongoing	

5 Summary

In this report, we described the first iteration of the Math-Bridge content release.We started by characterizing the source material initially provided for usage with the Math-Bridge service by the partners. We detailed the transformation procedure that was used to create this first content release and reported on the tools developed to assist the conversion process. We then stated the quality management approach we put to work and gave an overview on the current state of the Math-Bridge content.

A Content Authors

This appendix lists main authors and contributors of the Math-Bridgecontent collections

A.1 German Research Center for Artificial Intelligence

- Edgar Keßler
- PD Dr. Erica Melis
- Michael Dietrich

A.2 University of Kassel/University of Paderborn

- Prof. Dr. Rolf Biehler
- Prof. Dr. Bernd Billhardt
- Prof. Dr. Regina Bruder
- Prof. Dr. Reinhard Hochmuth
- Prof. Dr. Wolfram Koepf
- Prof. Dr. Walter Strampp
- Dr Reinhard Gerhold
- Pascal Fischer
- Juliane Klemm
- Isabell Kuhnke-Lerch
- Thomas Lange
- Alina Lompe
- Stefan Podworny
- Thomas Wassong

A.3 Open Universiteit Netherlands

- Evert van de Vrie
- Josje Lodder
- Bram van Asch
- Rene Beerends
- Paul Bezembinder
- Marlies Bitter
- Jan van de Craats
- Ton van Gijsel
- Anko Haven
- F. Martens
- Martijn Mulder
- Geya van Prooyen
- Maarten de Rijke
- F. Simons
- Elias Thijsse
- Jan Timmerman
- Heleen Verhage
- Benne de Weger
- Marcel Wigman
- Bert Zwaneveld

A.4 Saarland University

- Marianne Moormann
- Christian Groß
- Christian Bokhove
- Vásárhelyi Éva
- Rui Dong
- Verónica Guzmán
- Paul Libbrecht
- Hana Moraová

A.5 University of Vienna

- Franz Embacher
- Petra Oberhuemer
- Anita Dorfmayr
- Stefan Götz
- Michael Grosser
- Sascha Husa
- Maria Koth
- Natascha Riahi
- Markus Röthl

A.6 Tampere University of Technology

- Pasi Häkkinen
- Janne Kauhanen
- Markku Luhtalahti

- Sanna Mäki-Turja
- Olli Perttula
- Seppo Pohjolainen
- Kari Suomela
- Markus Urponen