



Internship report

Repository as supporting platform for the exchange of pedagogical material by students

Student: Carlos Gajardo Maureira [cgajardo@ing.uchile.cl]
Professor: J r my Barbay [jeremy.barbay@dcc.uchile.cl]

Spring 2011

Index

1. Introduction	2
2. Definitions	3
2.1. Crowdsourcing	3
2.2. Bootstrapping Databases.....	3
2.3. Repository.....	4
2.4. Sensei.....	4
2.5. U-Cursos	4
3. Objectives	5
4. Methodology	5
5. Preliminary Work	6
5.1. Participation model in Repository	6
5.2. Description of potential features.....	7
5.3. Description of expected result.....	8
6. Work	9
6.1. Survey results	9
6.2. Extending the data model.....	13
6.3. Other task	14
7. Pending Work	15
8. Conclusions	16
Appendix A: Survey details	17

1. Introduction

Supporting educational processes through computing is a constant work that requires continuous improvements and adaptations.

Group dynamics has changed; people interact in new ways, especially young and children, whom are more willing to participate and collaborate in communities to share interests or knowledge.

In this internship we propose the use of a repository in Repositorium as a platform for the exchange of pedagogical material by students, emulating the functionality of Sense project.

2. Definitions

The following are some general definitions required to understand the next sections.

2.1. Crowdsourcing

Traditional approach to solve extensive problems includes the assignation of tasks to specifically selected individuals. This approach is present in Internet too; for example, in online communities privileged users, the administrators, are in charge of control and moderate content.

Web 2.0 came to break some of these traditions, starting by the fact that now users are not only consumers of information, but also producers. Communities like YouTube relay on their users to create and moderate content.

Following Web 2.0 philosophy new models for solving problem have been proposed. Crowdsourcing is one of them. In this model problems are presented to an unspecific number of people or community, relaying on their abilities and knowledge.

It's important to mention that this model not only is applied in the web sphere, but also in real life problems. For example, people may be invited to gather resources and build tools in critical times. For instance, after the 2010 earthquake in Chile, a group of people was summoned through an open call to a CrisisCamp. The result of this work can be checked at <http://www.chileayuda.com>

2.2. Bootstrapping Databases

A Bootstrapping Database, BSDB, is a collective database that, using *Crowdsourcing* principles, tries to solve issues that every community must face, the lack of participation incentive and the control of contents' quality.

BSDB propose a participation model in which users are required to add content to the community or perform small tasks to control quality of the existing content, in order to gather access to new content. This is achieved by using a Quality Control Criteria where users must answer questions, or challenges, similar to the way ReCaptcha system works.

There is a developing implementation of BSDB principles named Repositorium.

2.3. Repositorium

Repositorium is a Bootstrapping Database engine that allows anybody to start a community for sharing and producing high quality textual documents. People can join the community and get access to the data by solving challenges and uploading new content.

It's an on progress open source project¹ developed by students at Departamento de Ciencias de la Computación in the Universidad de Chile with the guidance and supervision of professor Jérémy Barbay.

2.4. Sensei

Sensei is an online repository where users can share pedagogical material related to courses taught at FCFM², this material vary from reference guides to exam questions. The project has been online since 2003 and today gathers more than 157.000 documents.

2.5. U-Cursos

U-Cursos is the official FCFM web platform for communication between students and teachers. It offers courses, grades, schedules and calendars management and pedagogical material publication system. This platform has been online since 2001.

¹ Source code can be found at <http://github.com/jyby/repositorium>

² FCFM: Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile.

³ FCFM: Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile.
In CakePHP a behavior is a way to organize some of the functionality defined in models. It allows

3. Objectives

The main objective of this internship is to build an online community to share pedagogical material using Repositorium as supporting platform and Sensei data as a starting catalogue.

Other objectives would be:

- Implement a module for Repositorium to upload files.
- Test the performance of Repositorium, in order to decide if it is suitable as an alternative to Sensei, or not.
- Correct eventual bugs (in collaboration with other students) which may affect the project.
- Acquire knowledge about the implementation of Repositorium aiming to the development of a future Engineering thesis.

4. Methodology

The first step in this internship was to develop and apply a survey about current usage of Sensei. The objective behind this survey was to determinate the main disadvantages of this platform and its key features.

The next step was to extend the current data model of Repositorium to allow the attachment of files into documents. This was a key feature for this internship given the nature of the platform aimed to develop.

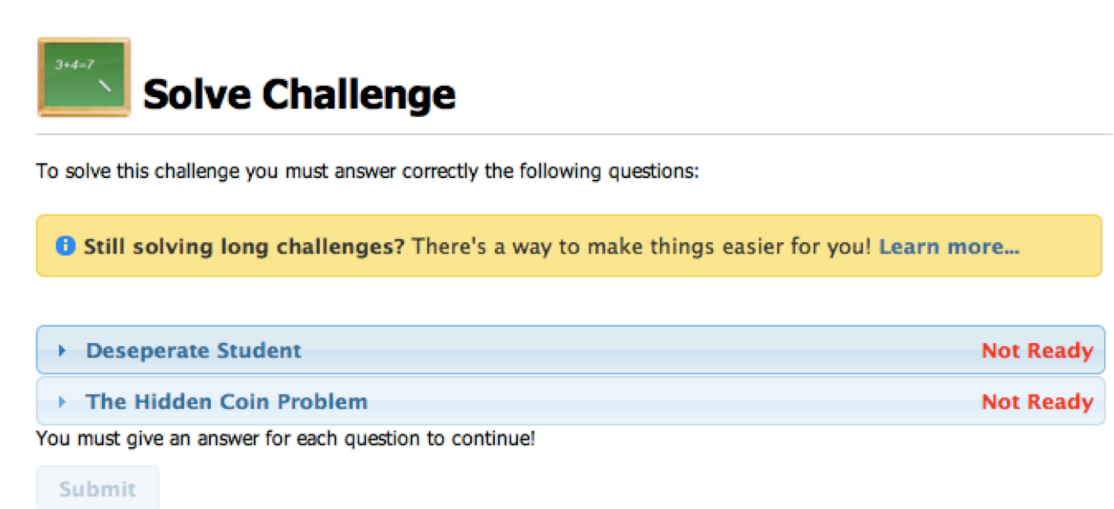
Coding, testing and integrating were the following step, which finished with the update of the current version of Repositorium running at <http://www.repositorium.cl>

A transversal activity was to detect and fix some bug in the system. The expected final step was to perform some test with final users.

5. Preliminary Work

5.1. Participation model in Repository

In Repository the members of each community are responsible for the quantity and quality of the documents. To incentive the participation Repository keeps a record of points for each user, this points can be earned by performing some quality control task or by uploading new documents to the repository. So everytime a user wants to download a document he has to spend some points, if the user doesn't have enough points then he will be forced to answers some questions or challenges, like the one showed in illustration 1.



Solve Challenge

To solve this challenge you must answer correctly the following questions:

Still solving long challenges? There's a way to make things easier for you! [Learn more...](#)

- ▶ **Deseperate Student** **Not Ready**
- ▶ **The Hidden Coin Problem** **Not Ready**

You must give an answer for each question to continue!

Submit

Illustration 1: a challenge example

In order to avoid malicious users who could give random answers to challenges, the system ask some challenges for which it knows the correct answer, so if the user gives an incorrect answer the system will keep track of it and will ask the user a larger number of challenges, making harder for the malicious users to acquire documents.

5.2. Description of potential features

The following are some potential features identified in the modelling phase.

F1. Adding restrictions on which elements compose the document

In the current version of Repositorium, documents are composed by a Title, some Content (text) and some Tags.

New features require some additional components, like attaching a File or URL, or additional restrictions, like limiting the Content size or restricting File extensions.

This new components and restrictions would be bound to Repositories at creation; nevertheless they can be changed later in section "Manage Repository". Some modifications to the model are required to achieve this, and those are shown in section "Extending the Model".

F2. Adding multiple criteria evaluation for a single document

In base version of Repositorium evaluation criteria are applied on a 1 to 1 basis, this means 1 criterion for 1 document, requiring some additional effort from users when files are part of a document, since this forces them to download as many files as documents are being evaluated.

In order to overpass this issue a new multi criteria evaluation is suggested. This would allow users to answer many questions (many criteria) about one document.

F3. Documents preview

Until now search results shows titles and text content, but F1 introduce new elements to documents, like files and url, this new elements must provide a preview of themselves.

Suggested previews are image of first page for PDF files, first ten seconds of audio or video files and low-resolution version of images.

5.3. Description of excepted result

The main objective of this internship is to build an online community for sharing pedagogical material between students. This includes sharing quizzes and exams questions from past semesters, but also slides, summaries, mind mapping diagrams, etc. These documents are expected to be from simple text files to images, PDF or Microsoft Word files.

New rules will apply within the Repository context, especially those regarding the management of documents:

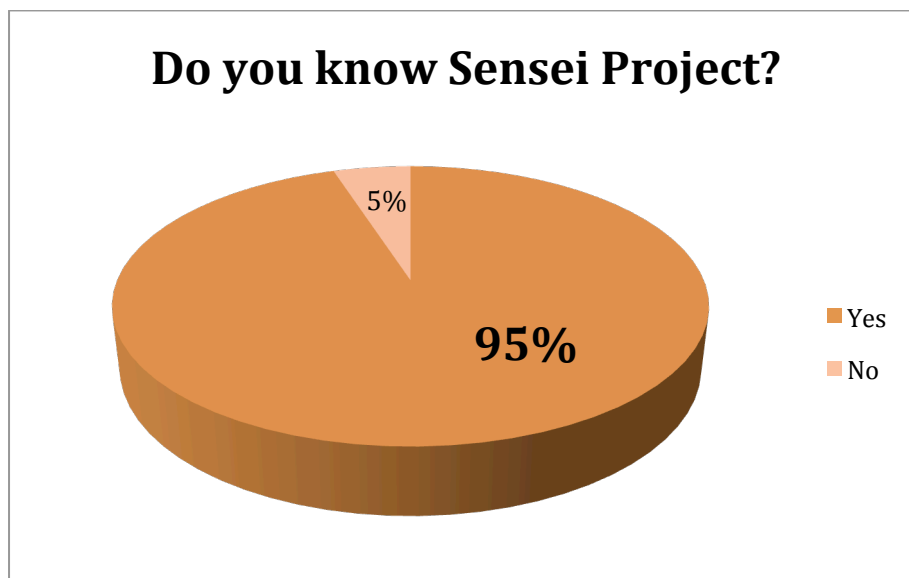
- In this platform the *Title* of a document will be used to describe and name the document attached. Examples of valid titles are: "Autumn 2010 Exam", "Summary for Quiz 1". Invalid names are: "Summary", "Quiz".
- The *Content* field will be used to add some additional description to the document, for example: "This exam took 3 hours and students were allowed to bring notes and books"
- Tags will be used to specify the department and course for which the document is related to. For example, for the Physics I Exam valid tags would be: "Physics I", "P101" and "Physics Department".
- Every document will allow a maximum of 3 files to be attached. If a document requires more than 3 files, the recommendation will be to split the document in two or more new documents. Usually a document will require no more than 2 files, the statement and its solution, but in some cases, if the files are photos of the statements for example and the document is composed of more than 3 statements, the recommended way is to create a new document for each statement, for example: "Quiz I Statement I", "Quiz I Statement II".

6. Work

6.1. Survey results

The main objective of this survey was to estimate the current level of usage and participation in Sensei. This survey was conducted between September 22th and September 23th of 2011 using Google Form and was answered by 430 students of FCFM. A full list of the questions can be found in Appendix A.

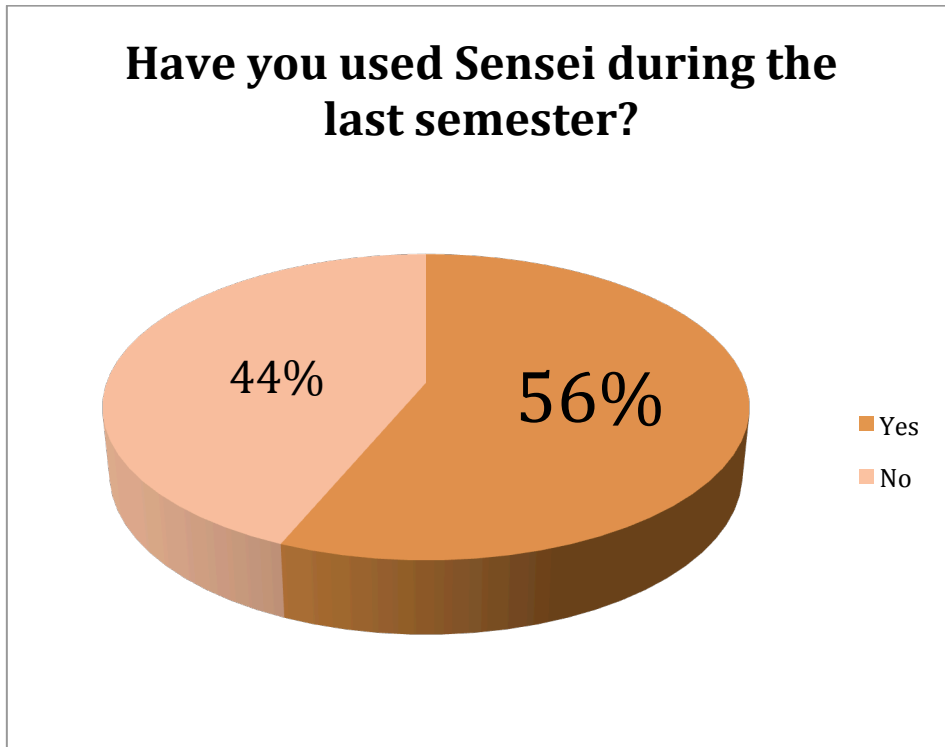
The first question tries to determinate how known is Sensei among students.



Graph 1: Do you know Sensei?

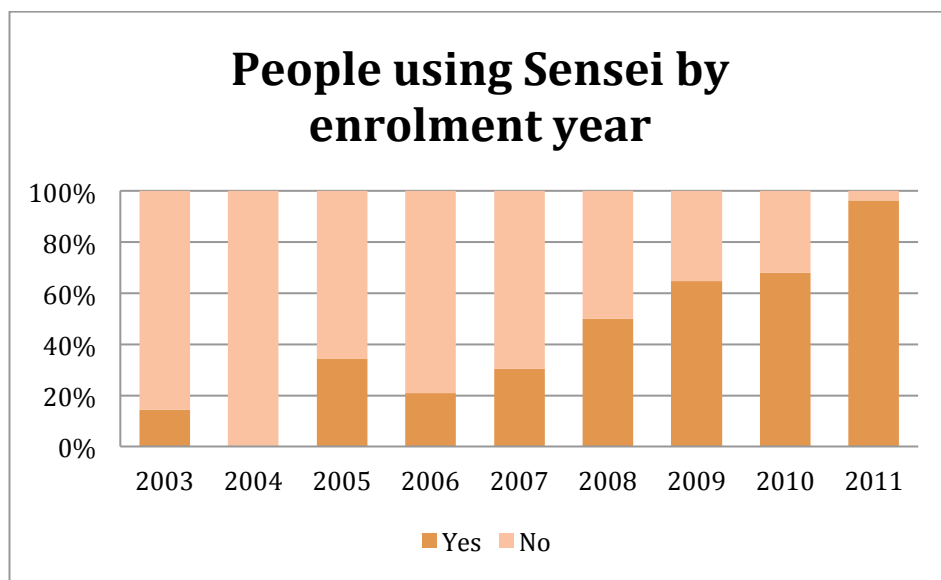
As seen in Graph 1, 95% of current students know Sensei project.

The second questions tries to indicate the level of usage. Students were asked if they have used Sensei during the last semester (Autumn 2011).



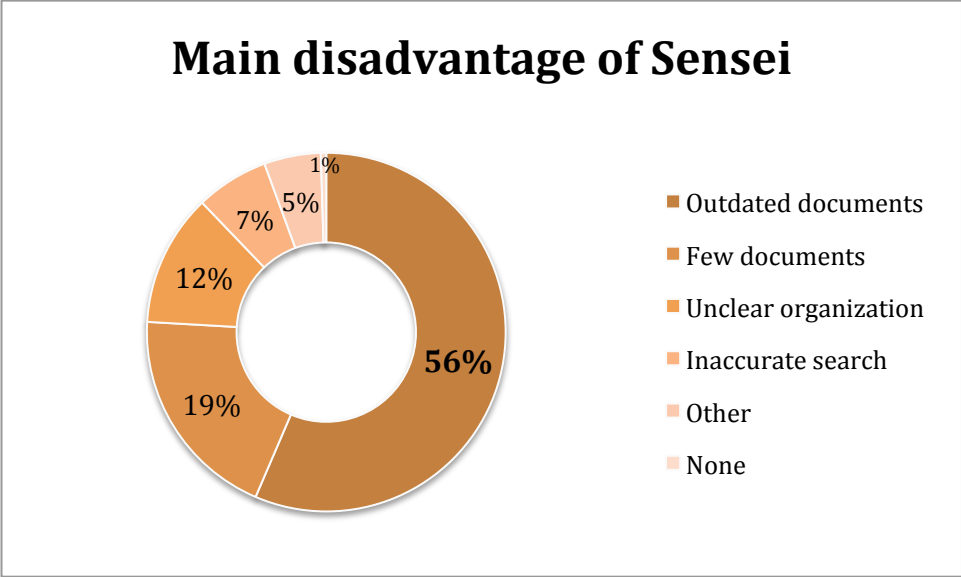
Graph 2: have you used Sensei during the last semester?

Graph 2 shows that, even when Sensei is known among 95% of current students, it's used only by the 56% of them. The next graph breaks down the data of Graph 2 by enrolment year.



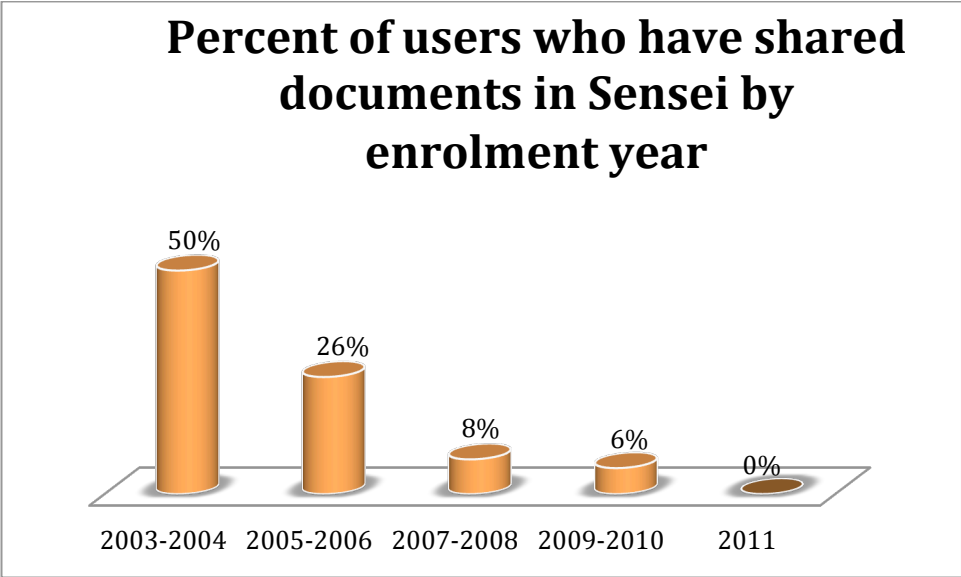
Graph 3: students using Sensei by enrolment year

It's clear in Graph 3 that Sensei is mainly used during the first 4 years of study. Students were asked then about the main disadvantages in current version of Sensei; the results are shown in Graph 4.



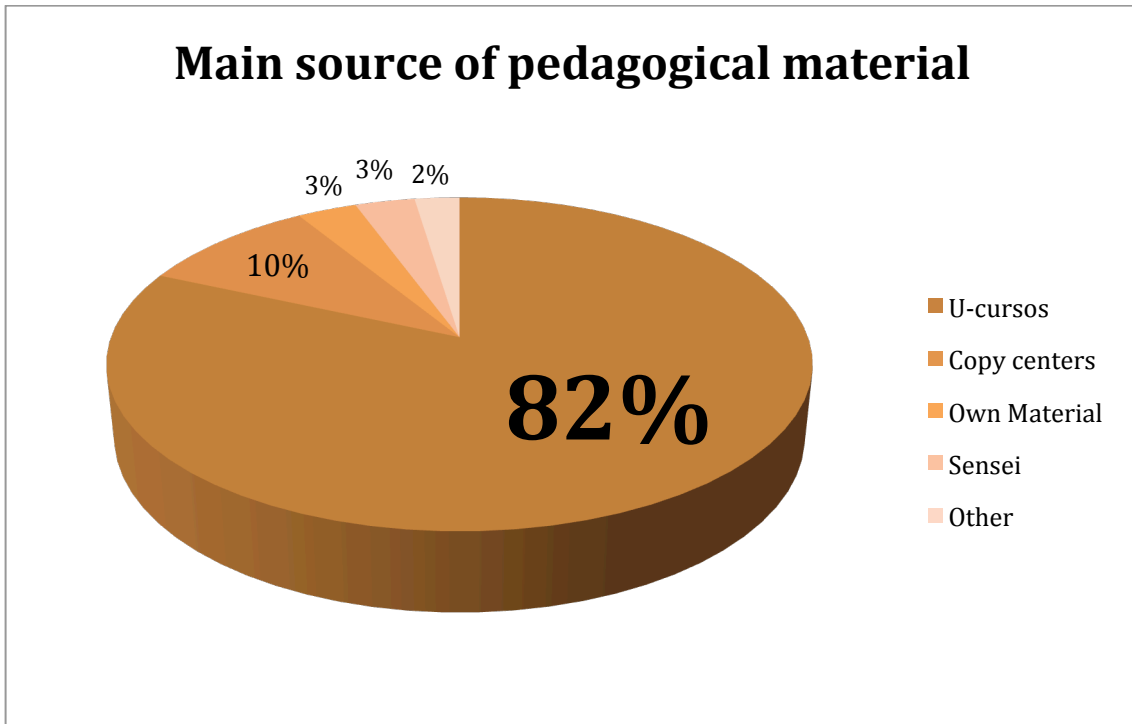
Graph 4: Main disadvantage of Sensei

“Outdated documents” was the main disadvantage of Sensei according to current users. This can be explained by the lack of participation in sharing documents; students were asked if they have ever shared a document in Sensei, the result are broken down by enrolment year and show in Graph 5.



Graph 5: percentage of users who have shared documents in Sensei by enrolment year

Finally students were asked about their main source for pedagogical material. It’s clear, as shown below, that Sensei is among the less used source and U-Cursos is the preferred place to search for pedagogical material.



Graph 6: main source of pedagogical material

In conclusion the survey shows that even when most of the users know about Sensei only 56% have used it during the last semester. It's clear then that Sensei is lacking some important feature or that users are expecting something more.

According to this survey what users are expecting are variety of contents, updated documents and a clear organization. This explains why users prefer U-Cursos as the main source for pedagogical material, given that they find up-to-date documents and a simple organization by semester and year; even when U-Cursos isn't designed to offer easy access to the documents; It's common to find courses where material is inaccessible due to the restrictions.

In Repositorium the variety of content and the updated documents will depend on whether users participate and collaborate or not. The organization will be ensured if the users adhere to the rules mentioned in section 5.3. As for the search of material, Repositorium offers a very simple and clear search engine with autocompleting and the option to select some quality criteria. All these features make Repositorium an excellent candidate to replace Sensei.

6.2. Extending the data model

To satisfy feature request F1 (Adding restriction on which elements compose the document) new entities are required, “Constituents” and “Restrictions”.

A Constituent represents one of the elements composing a Document. *Name* field stands for the Constituent name presented to users; *description* field describes the Constituent, this description is also presented to users; *sysname* field indicates the internal name of the Constituent, and is used by the system to locate view elements, *sysnames* with no whitespaces are preferable in order to avoid file system issues.

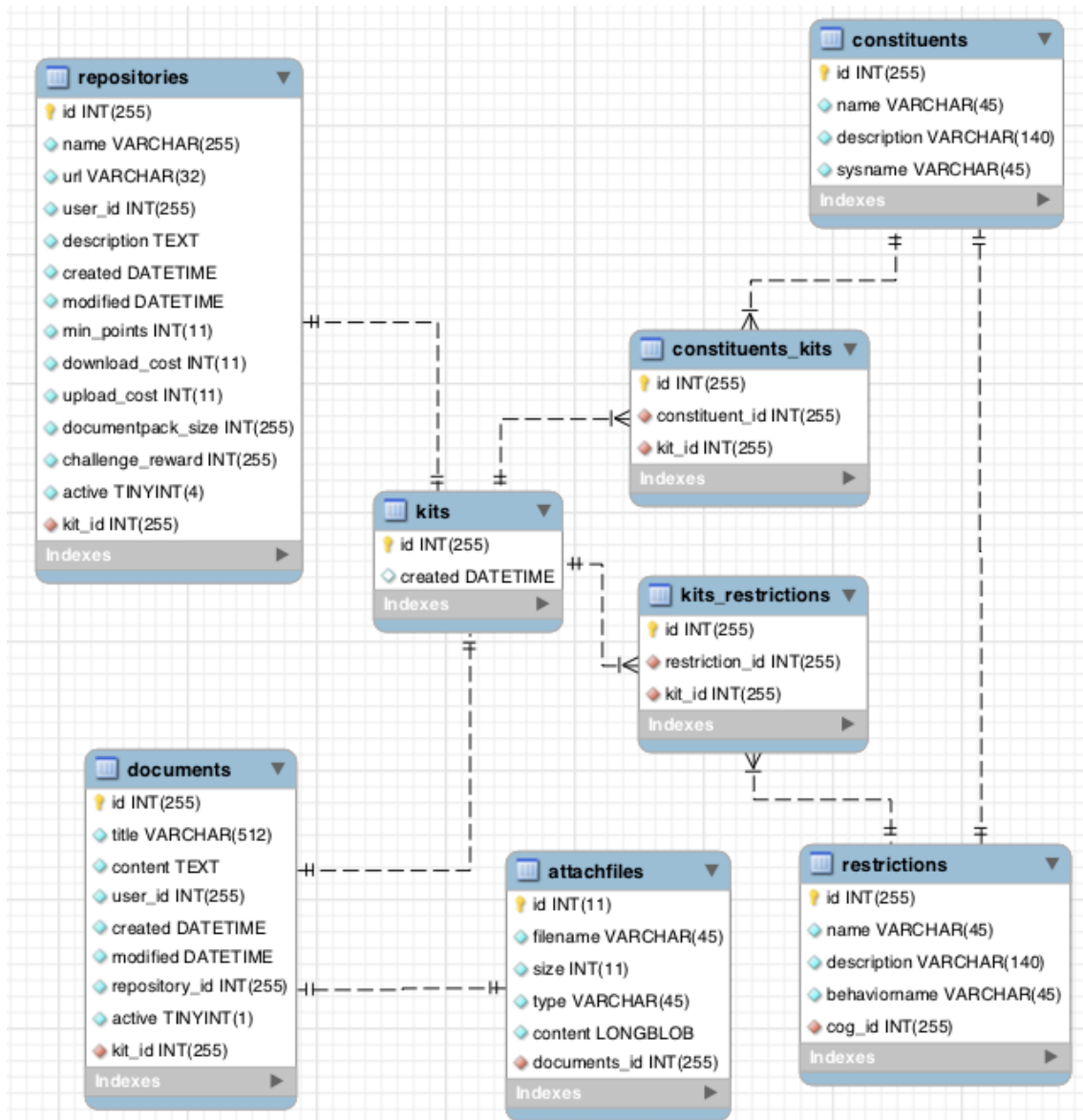


Illustration 2: Data model of Repository

A Restriction represents a rule to be applied when a Document is created, allowing the Repository manager to have a sharper control over the type of content user can upload. Every Restriction is associated to a Constituent. *Name* field stands for the Restriction name presented to users; *description* field describes the Restriction, this field is also presented to users; *behaviorname* field indicates the system name for the behavior³ to be applied.

Constituents and Restrictions are related to Repositories through Kits. A Kit is a selection of Constituents and Restrictions defined by user when the Repository is created. This selection can be changed later in the section “Manage Repository”.

Documents are created with the Kit defined in the container Repository. Whenever this selection changes a new Kit is created and *kit_id* is updated in the Repository, so the Documents already created stay linked to their original Kit.

The first Constituent to be developed is “Attach File”; this Constituent will allow users to add files (of any type) to a Document. This new Constituent requires the Attachfile entity to work. Attachfile entity represents a file attached to Documents when “Attach File” Constituent is activated in a Repository

This design was conceived keeping two principles in mind, that the implementation of the model shouldn’t interface with the rest of the system (encapsulation), and that it should be easy for future developers to add new constituents or restrictions (extensibility).

6.3. Other task

Other tasks performed along the semester were:

- The development of a script to gather files from U-Cursos.
- The development of a script to automatically upload Latex files to Repositorium.
- Some issues were fixed related to the deletion of documents in Repositorium.
- The update of the current version of Repositorium running in Anakena server.
- The implementation of the feature to show files and tags in challenges. This will allow performing quality control over files and tags.

³ In CakePHP a behavior is a way to organize some of the functionality defined in models. It allows separating logic that may not be directly related to a model, but needs to be there.

7. Pending Work

Even when Repositorium has been updated to meet the technical requirements for a sharing pedagogical material community there're still some tasks pending related to:

- Populate the system with an initial catalogue of documents.
- Open access to a group of final users to gather opinions about usability and possible new features.
- Integrate the development completed by the student of Software Engineering II. This work includes some important features, like extending the search engine to include not only tags but also title and content and the detection of duplicated documents.
- Make a second update to the currently running version of Repositorium.

8. Conclusions

The main objective of this internship was to build an online community to share pedagogical material using Repositorium as supporting platform. Some work was completed in order to allow Repositorium to attach files to every document and to prepare the platform for this type of communities, but there's still some work pending in order to achieve this goal; mainly to open the access to users so this software can become a real community. It's clear that there's a need for a platform like this and users showed big interest in this project.

Some bugs were fixed and some features were added, like the capability to show files and tags in challenges. This is an important feature and must be added to the currently running version of Repositorium prior the opening of the platform to final users.

Once the pending tasks are completed Repositorium will be in a good position to become the platform for the community of pedagogical material sharing.

Finally all the study and work of the platform gave me a good insight of the system and will allowed me to perform better when working for my Engineering thesis.

Appendix A: Survey details

Name: ¿Conoces Sensei?

Objective: measure the current level of usage of Sensei.

Start date: 22/09/2011 20:00

End date: 22/09/2011 20:00

Questions

¿Conoces el proyecto Sensei?

Objective: Check how known is the project.

Choices: [sí | no]

¿Haz usado Sensei durante el último semestre?

Objective: determinate the overall level of usage of Sensei.

Choices: [sí | no]

¿Cuáles de las siguientes características haz usado?

Objective: determinate the level of usage of community features.

Choices: [Comentarios | Votaciones | Ninguna | Otras]

¿Haz compartido material en Sensei?

Objective: to know users' participation level

Choices: [sí | no]

¿Cuál crees que es el mayor defecto de Sensei?

Objective: identify the reason why Sensei is less used every year.

Choices: [material desactualizado, organización de los archivo, poco material, es difícil subir archivos, es difícil encontrar lo que se busca, otro]

Quando buscas material de estudio, ¿cuál es tu primera fuente?

Objective: Identify sources for pedagogical material.

Choices: [sensei | u-ucursos | fotocopiadora | material propio | otro]

¿En qué año ingresaste a Beauchef?

Objective: separate pre and post u-cursos students.

Choices: [2003 , 2011]