

SliDL: A Slide Digital Library Supporting Content Reuse in Presentations*

José H. Canós, María Isabel Marante, and Manuel Llavador

Dept. of Computer Science (DSIC), Technical University of Valencia,
Camino de Vera s/n, E46022, Valencia, Spain
{jhcanos,mmarante,mllavador}@dsic.upv.es

Abstract. Presentation building applications lack good support to slide reuse. In this paper, we introduce SliDL, a digital library that facilitates slide reuse by flattening the presentation-based structure of current systems and providing slide retrieval facilities. The service-oriented architecture of SliDL enables slide sharing between different applications. We have developed clients for Microsoft PowerPoint 2007 and OpenOffice.org Impress.

Keywords: Slide reuse, presentation management, Service-Oriented Architecture.

1 Introduction and Motivation

Content reuse is one of the unsolved matters of current presentation tools. Although systems like SlideShare¹ allow sharing presentations, their presentation-centered approach makes it difficult to reuse slides belonging to other presentations during the preparation of a new one. To reuse a slide, a user must know the presentation it belongs to, open the presentation, look for the slide, and copy and paste it when found. This becomes a serious drawback when the amount of presentations is large. Only the newest version of Microsoft PowerPoint includes a feature to support slide reuse: Slide Libraries² can be used to export slides, which can be included in further presentations. This feature requires the computer to be connected to a server running Microsoft Office SharePoint Server 2007, and in this case, only slides created with PowerPoint can be reused. Consequently, a PowerPoint user cannot find nor import slides created with other programs such as OpenOffice.org Impress. Moreover, slides can only be found via browsing the entire server library, without support to neither content-based nor metadata-based searches on clients.

* SliDL has been implemented by students of the 2008-2009 edition of the Digital Libraries course of the “Master en Ingeniería del Software, Métodos Formales y Sistemas de Información” at the Universidad Politécnica de Valencia. The work of J. H. Canós and M. Llavador is partially funded by the Spanish Ministerio de Educación y Ciencia (MEC) under grants META (TIN2006-15175-C05-01) and CAPES/DGU 2008 (PHB2007-0064-PC), and Generalitat Valenciana (ACOMP07/216). M. Llavador is the holder of the MEC-FPU grant #AP2005-3356.

¹ www.slideshare.net

² <http://office.microsoft.com/en-us/sharepointserver/HA101741171033.aspx#1>

SliDL is a Slide Digital Library developed to reduce the difficulties of slide reuse. The overall goal of SliDL is to help users to build presentations by reusing previously generated slides contained in some presentation files. This means that it must provide services for storing and retrieving slides. The following features are included in the current release:

1. Platform independence: users are able to store presentations and reuse slides generated with different programs. A platform independent slide repository stores slides created with different applications.
2. Cross-platform usage: SliDL has been implemented following a service-oriented approach, so that the storage and retrieval services can be shared by every client application.
3. Slide metadata management: some slide properties are stored to allow metadata-based retrieval.
4. Content-based slide retrieval: the textual content of slides is indexed to allow keyword-based slide search.
5. Slide collection browsing: SliDL users can browse the thumbnail collection to find slides. Browsing can be made over all the slide collection or over the results of a previous content and/or metadata based search.

We describe the architecture of SliDL and the services provided in its current release. We also outline some of the distinguished features of the forthcoming release.

2 The Service-Oriented Architecture of SliDL

Figure 1 shows the layered architecture of SliDL. At the Storage Layer, the SliDL Repository stores the slide collection. The repository is updated and queried through Indexing and Searching components of the SliDL API at the Service Layer. Presentation software at the client layer will use the services of the SliDL API to send and retrieve slides information to and from the repositories. We give a more detailed description of each layer in the following subsections.

2.1 Client Layer

The Client Layer is where the presentation software is placed in SliDL. In order to be able to interact with the SliDL services, clients for the different programs must be developed. A client is a small program able to handle a presentation using the standard application's object model (i.e., a plug-in for the specific application). A SliDL client has two main components (see Fig 1). On one hand, the *Slide Extractor* processes a presentation and extracts its slides; then, it uses the services of the SliDL API to send the slides, along with their associated metadata, to the repository.

On the other hand, slide retrieval facilities are provided by the *Slide Seeker*. The module offers several ways to retrieve slides:

- Metadata-based search: typical searches based on author, title, date, etc.
- Content-based text retrieval: the user can type one or more words, and the slide seeker returns a list of the slides containing such words in some of their textual shapes.

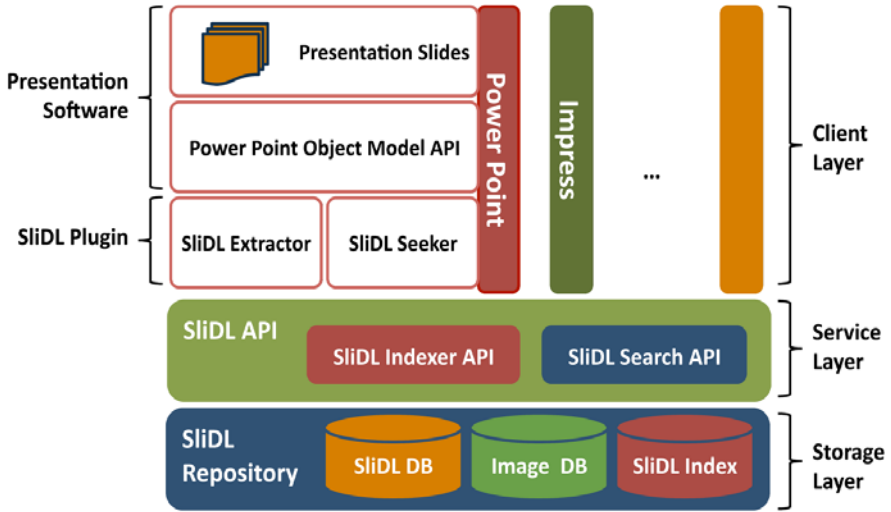


Fig. 1. Architecture of SliDL with details about SliDL clients

- Thumbnail-based slide browsing: the user may select the requested slide clicking on the corresponding slide.
- Mixed strategies: a user can first make a content-based query, and then browse the thumbnails of the returned slides to precisely locate the slide sought.

2.2 Service Layer

The service layer offers services for storing and retrieving slides from the SliDL repository, which is accessed via a service-oriented API (<http://mborges.dsic.upv.es:8001/SLIWS/service.asmx>). To provide interoperability, we have defined a common slide exchange format. The slides sent/obtained by clients to/from the repository are encoded in XML.

2.3 Storage Layer

The SliDL Repository holds the collection of slides; for each slide, the following information is held:

- Descriptive and structural metadata coming from both the slide and the presentation it belongs to (e.g. author, title of presentation, title of slide); slides metadata are stored in the SliDL DB component.
- Text content: all the text contained in textual shapes, including headers and footers.
- Thumbnail: a small image of the slide. The purpose of thumbnails is twofold. On one hand, they can be used to browse through the slide collection; on the other hand, they can be used in the future to use some image-based slide retrieval facilities.

The process of storing a presentation in the Repository is as follows: the SliDL Extractor of a client processes a presentation every time it is saved after some change³. For each slide in the presentation, index terms are extracted from the content of text shapes, and put into the SliDL Index; in addition, a thumbnail of the slide is generated and added to the Image DB; finally, metadata is extracted from the slide and the container presentation and inserted into the SliDL DB.

3 Current Status and Further Work

The current version of SliDL is for personal use, that is, without support to multiple users. Its goal is to help users to organize their presentations. We have developed clients for both Microsoft PowerPoint 2007 and OpenOffice.org Impress.

We are moving to a new version where the following features will be available:

- Multi-user facilities: SliDL will allow users to define a personal space within the digital library and decide whether the content of such space must be considered private or accessible to other users of the system.
- Slide tagging and rating to enhance the retrieval facilities.
- Tracking and reviewing changes to slides on the server to keep user's presentations updated.
- Duplicate slide detection to find slides included in different presentations.
- Metadata-based recommender systems: exploiting relationships such as "belong to the same presentation" may enrich the search results; also, as a mid-term goal, it would be desirable to build systems able to suggest a preliminary presentation from a set of keywords using the slides contained in SliDL.

³ It is possible to explore folders and process all presentation files found as an iteration of the described presentation indexing process.