

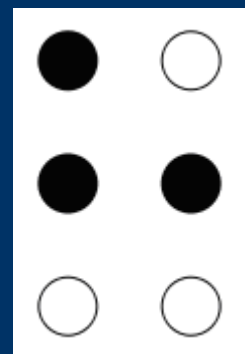
Translation of PDF and ODF documents to Braille

Ian Ball
12413959

Supervisor: Iain Murray

What is Braille?

- Braille is a system of communicating, primarily for use by blind people.
- Each Braille character consists of dots in a 2x3 matrix.
- For example, the character “h” would be as shown.
- There is also “Grade 2” Braille, which is similar to shorthand.



What is PDF?

- Portable Document Format is a document format developed by Adobe, that is designed to represent documents in a platform independent manner.
 - A de-facto standard for distributing documents.
 - Included with Mac OS X 10.4, Apple has provided a set of libraries called “PDF Kit” that allows PDF documents to be manipulated, and for the contents to be extracted.
-
-

What is ODF?

- OASIS Open Document Format is a document format for representing text documents, spreadsheets, presentations, graphics and mathematical formulae.
 - XML file format, with several XML files inside a JAR archive.
 - ISO Standardised
-
-

Computerised Braille Translation

- While there are many products that currently exist for computerised Braille translation, they are very expensive.
 - There is a lack of software that translates PDF or ODF documents to Braille.
 - There is no other software that allows Braille conversion on the Mac OS X platform.
-
-

Computerised Braille Translation

- There is an online service called RoboBraille, that can convert text documents and Microsoft Word documents to Braille free of charge. However, it is not possible to edit the document after it has been converted.
 - Most commercial packages do not allow “on-the-fly” conversion, where the text can be edited and the conversion performed immediately
-
-

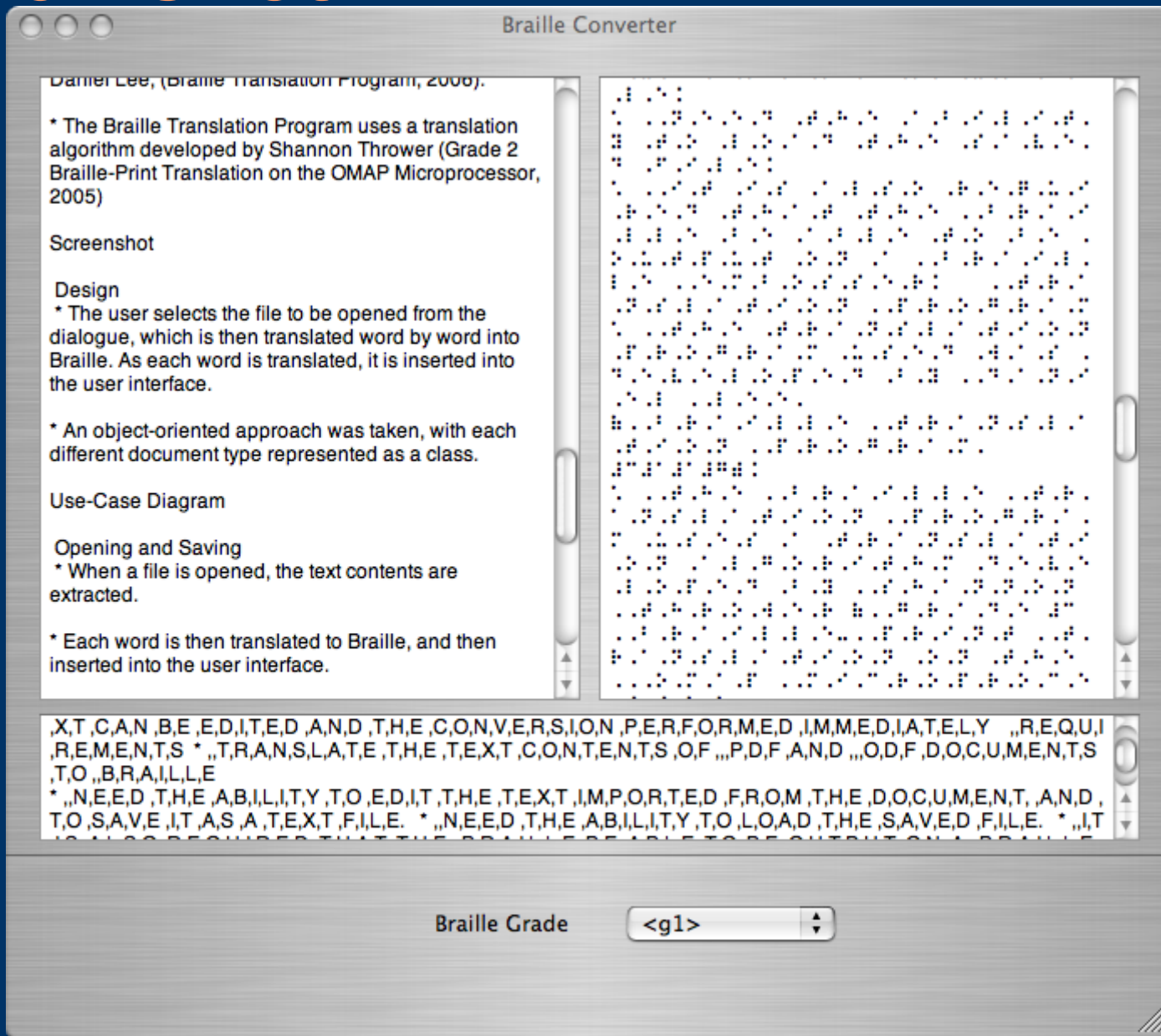
Requirements

- Translate the text contents of PDF and ODF documents to Braille
 - Need the ability to edit the text imported from the document, and to save it as a text file.
 - Need the ability to load the saved file.
 - It is also required that the Braille be able to be output on a Braille Embosser.
-
-

Translation Program

- The translation program used was developed by Daniel Lee, (Braille Translation Program, 2006).
- The Braille Translation Program uses a translation algorithm developed by Shannon Thrower (Grade 2 Braille-Print Translation on the OMAP Microprocessor, 2005)

Screen shot



Design

- The user selects the file to be opened from the “Open File” dialogue, which is then translated word by word into Braille. As each word is translated, it is inserted into the user interface.
 - An object-oriented approach was taken, with each different document type represented as a class.
-
-

Use-Case Diagram



Opening and Saving

- When a file is opened, the text contents are extracted from the file.
 - Each word is then translated one at a time to Braille, and then inserted into the user interface.
 - When the text is saved, two files are created. One contains the plain text representation, and the other contains the ASCII Braille representation.
-
-

Supported Formats

- The translation program will convert files of the following types
 - .pdf (Portable Document Format)
 - .odt (Open Document Text)
 - .odp (Open Document Presentation)
 - .txt (Plain Text)
 - In addition, plain text files such as programming code (.c etc) can be translated.
-
-

Document Formatting

- The translation works best on documents with a simple structure.
 - Formatting of the document is not able to be preserved (for example: bold, underline, centring, etc).
 - Headers and Footers should be avoided, since the text from these will not be distinguished from the body text.
 - Graphics are obviously not translated :)
-
-

Implementation

- Implemented using Objective C using Xcode on Mac OS X 10.4.
 - PDF text extraction was done using the PDF Kit, supplied as part of the Cocoa libraries
 - XML Parsing done using the XML Parsing classes supplied as a part of Cocoa
-
-

Future Developments

- Add additional document formats, such as Microsoft's OpenXML.
- Improve the support for Open Document Format Documents, by increasing the number of XML elements recognised.



Questions ?

