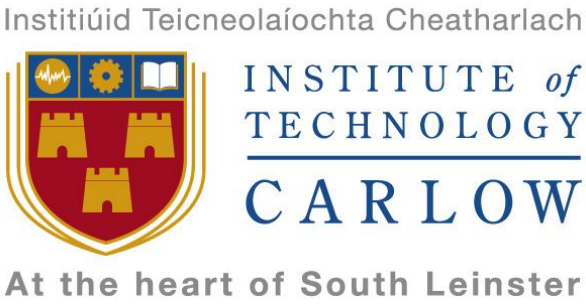
Research Document

ExamIT

AUTOMATIC TEST CORRECTION PLATFORM

4th year, Software Development project

Institute Of Technology Carlow



Roger Marciniak

Student Number: C00169733

Supervisor: Paul Barry

31st October, 2016

Table Of Contents

Introduction

The purpose of this document is to discover and evaluate the available technologies and solutions in order to introduce full functionality to the system, as specified in the Functional Specification Document.

The research document will cover the already existing solutions, with functionality similar to the specification of this project. Apart from that, the available database technologies will also be evaluated, and the most fitting one will be selected.

Much attention will be given to vision libraries that fulfill the project specification. Vision library choice will be the most serious choice of this project and the whole correction algorithm will depend on the library quality, speed, ease of use, documentation quality and adaptability. The library choice will also affect the language in which the project will be written, depending on which languages the library is available for.

Lastly, the backend solution will have to be selected that caters to the language, libraries and database chosen.

Existing Solutions

Exam Professor

Exam Professor is an online service allowing for the creation of multiple choice tests. These are online tests which a user can create after setting up an account. The time for completion of the test can be specified. In case of an internet outage, the students are able to pick up where they have left off because the partial results are saved each step of the way.

There are anti cheating measures like randomised question order and selecting an amount of questions from a larger pool of questions, therefore the each student will receive a slightly different test.

Through a private login, student activity can be tracked and later scored, available to view and printed or exported. An amount of retakes can also be specified. The user also chooses what the students see after their exam is graded. [E1]

The user can choose to receive automatic notifications.

Other functionality includes:

- Attaching images
- Question statistics
- Export to PDF, Excel files
- Exam can be embedded on any site

Exam Professor is a paid service, allowing for fourteen days of free trial. Later, the user has to buy one of their plans at $24, $49 and $99 per month. [E2]

Quizworks

Quizworks is an online quiz creation tool with multiple choice test functionality. There are six types of questions to select from, these include image questions, boolean questions, long eight answer questions, series of images and variations of those already listed.

The user is able to customise the look and feel of the test, and personalise social media settings. Time for completion can be set and the number of available redos. Questions can be sorted into categories and information can be a mandatory field or optional. The users taking the test can receive automatic PDF certification. [E3]

Additional functionality includes:

- The tests can be embedded on the users website
- Advanced user management (very vaguely described)
- Rich statistic about the user levels

Quizworks is a paid service with plans for $20, $50 and $99 per month. Where the $50 plan offers all the above functionality and is the cheapest one that has support. There is a free version but is very limited and would not suffice for the needs of this project. [E4]

Easy Test Maker

Easy Test Maker is an online test generator for the creation and management of online tests. It allows for multiple choice, fill in the blank, matching, short answer and boolean questions. The user can add dynamic instructions and divide the test into sections.

The tests are graded automatically and can be viewed or printed, the user can also override grading when necessary. Also, reports are generated so that the user can see which questions caused most trouble.

The test can have dynamically created alternate versions of different question order to reduce cheating. Tests can be edited and reused. [E5]

Additional features include:

- Exporting to PDF or Word document
- Auto formatting
- Question dragging and reordering

Easy Test Maker is also a paid service with a very basic featureless Free plan, a $44.99 per year Plus plan with most of the features but a 250 limit on test results and a Premium plan which costs $74.99 per year. The last plan adds extra functionality like images, charts, graphs. [E5]

Conclusion

The project is based on the idea that the testing platform should be web based so that it does not to be installed and can be used from any device anywhere, therefore web based services were evaluated for the purpose of this document.

However, none of these services actually utilise vision software to correct scans of paper based tests. All of them are online, web tests. This was not the perfect situation as there were no similar solutions already available to help develop the project.

Another problem is that the project is aimed at use inside the institute, therefore it would be best if the user data and the system were installed locally, inside the institute. None of the available solutions cater for that. None of the solutions can be customised for the needs of the college

Other solutions like those provided by Xerox [X1] are extremely expensive and are used by state examiners across the world and their features are hard to find. The system also requires specialised scanner machines. The assumption is that versions of the software are created per customer.

Overall, the online solutions are either too simple or do not suit the specification of the project. Because of this, many features of the project will have to be designed from scratch without the ability to model on existing strategies.

Database Technology

The database choice for any project comes down to the choice between an SQL and NoSQL database. The research took a single article to realise that a NoSQL database was needed. [D1]

This is because:

1. The data will be ‘unstructured’
2. Some fields will contain lists of items instead of single values eg. list of answers to a question.
3. NoSQL databases scale horizontally, which means that all that needs to be done to scale the database is to add resources in the cloud or partition the database over more servers.
4. If the data was forced into a tabular design, to display a test it would be necessary to perform multiple joins. This is very uneffective and resource consuming.

NoSQL Choice

There are many NoSQL solutions using a key/value structure, column, document and graph structures. The document format is the most suitable for storing the data of this project.

The document structure is very similar to column but allows for more nesting. Having answers inside a question inside a test with multiple questions fits the document structure perfectly. The document solution overcomes the constraints of the one or two level nesting constraint in the columnar solution. Imagine a document inside a document and so on.

The downside is that fetching a value in a document means getting the whole document, same goes for updates. All of this affects the performance, but still is the most suitable out of the available options. [D2]

Document based databases are the most popular out of the NoSQL family, and out of the document based ones two stand out and are used most often: MongoDB and Couchbase. Of the two, MongoDB is the more popular one.

- MongoDB is ranked #5 overall database and Couchbase #23 according to DB-engines Ranking chart [D3]
- Both are open source and run on Linux, Windows, OSX
- MongoDB supports twice as many languages (14 vs 27)
- Both use JavaScript for server-side scripting

Either of the technologies would work but the author is already familiar with MongoDB, therefore it was chosen as the database technology for the project.

Vision Library

When researching Vision libraries OpenCV stands out, being mentioned nearly in every internet source. Three of the most popular are Matlab, OpenCV and SimpleCV.

Matlab is not just a library, it is a complete scientific suite which consists of a large IDE and its own language. For this project, this is not an ideal solution. Firstly, learning a new language would not be optimal as there is a limited time frame to complete the project. Secondly, the project is supposed to be as efficient as possible, there is no need for a complete suite, and what is needed is a specialised, simple tool to perform the test correction. [L1]

Therefore the choice is between OpenCV and SimpleCV.

- Both are completely open source
- OpenCV is vastly more popular
- OpenCV is used by: Google, Yahoo, Microsoft, Intel, IBM, Sony, Honda, Toyota [L2]
    - Google stitches streetview images with OpenCV
    - Used for surveillance
    - Helps robots navigate and pick up objects
    - Supports C++, C, Python, Java, Matlab and runs on Windows, Linux and OSX
    - Best vision library documentation
- SimpleCV is a framework providing access to many vision libraries [L3]
    - Comes with an interactive shell for testing
    - Great for quick small projects
    - Designed for Python
    - Simpler than OpenCV
    - Poorer documentation

SimpleCV fits ideally into ‘the simple tool to perform the test correction’ idea. The fact that it partially uses OpenCV under the hood means that the user can still get the speed benefit of a library written in C++.

On the other hand, OpenCV has better documentation and quickly comparing the two shows that the problems encountered during the writing of the algorithm would be easier to solve with the help of the vast OpenCV documentation. Because this is the author's introduction to vision software it seems reasonable to use a well documented library.

Other benefits include:

- Not limiting development to one language
- Many tutorials and examples online
- More popular library means finding help online is easier
- Non-project benefits:
    - Learning the most popular vision library might pay off in the future
    - Learning a library that has implementations in many languages might also pay off in the future

Therefore, OpenCV was selected as the library for the purpose of creating the correction algorithm.

Development Language

Since OpenCV, the library of choice supports C++, C, Python, Java and Matlab, the author now has to select a development language. [L2]

It was agreed before that learning Matlab would consume too much project time. Left with four choices the author decided to develop the project using Python.

This was because:

- C++ is has the fastest, most efficient implementation and the Python implementation utilises the C++ code, it is a wrapper for the C++ implementation
- Writing in Python will be quicker
- Best of both worlds
- Python has great web frameworks that can be used for the project interface and backend.

Backend

Since the institute uses Linux servers and the author is most proficient using Linux it is reasonable to choose Linux as the server running the project backend. This will simplify the installation process on the institute's server and allow for quick deployment.

Since, Linux uses a package system instead of browsing the web for the necessary software, the author can install them using the package manager. This is a much simpler process for an experienced Linux user than downloading .exe files for Windows.

Because, Python is the language chosen for development, it is necessary to select a Python web framework. The two most popular frameworks are Django and Flask. Django is more feature rich and many large, popular services like NASA, Instagram and Pinterest run on it. [L4] However, for the purpose of the project a framework is required as simple as possible to be able to focus on writing an effective and efficient correction algorithm. Also, the author is already proficient with Flask [L5], this will further speed up the development process.

Therefore, Flask was chosen as the web framework to build the service with.

References

[E1] Exam Professor. (2016), Tour Page, [online], available:
http://www.examprofessor.com/tour
[accessed 22 November, 2016]

[E2] Exam Professor. (2016), Pricing Page, [online], available:
http://www.examprofessor.com/prices
[accessed 22 November, 2016]

[E3] Quizworks. (2016), Products Page, [online], available:
https://www.onlinequizcreator.com/products/item243
[accessed 22 November, 2016]

[E4] Quizworks. (2016), Pricing Page, [online], available:
http://pricing.onlinequizcreator.com/
[accessed 22 November, 2016]

[E5] Easy Test Maker. (2016), Start Page, [online], available:
https://www.easytestmaker.com/
[accessed 22 November, 2016]

[X1] Xerox (2016), Xerox Test Scanner Manual, [online], available:
http://www.xeroxscanners.com/downloads/manuals/DM4790/DM4790_Remark.US.pdf
[accessed 23 November, 2016]

[D1] Microsoft Azure (2016), NoSQL vs SQL, [online], available:
https://docs.microsoft.com/en-us/azure/documentdb/documentdb-nosql-vs-sql
[accessed 23 November, 2016]

[D2] Digital Ocean (2016), A Comparison Of NoSQL Database Management Systems And Models, [online], available:
https://www.digitalocean.com/community/tutorials/a-comparison-of-nosql-database-management-systems-and-models
[accessed 23 November, 2016]

[D3] DB-Engines (2016), System Properties Comparison Couchbase vs. MongoDB, [online], available:
http://db-engines.com/en/system/Couchbase%3BMongoDB
[accessed 23 November, 2016]

[L1] University Of Wisconsin, Space Science and Engineering Center (2016), What is Matlab, [online], available:
http://cimss.ssec.wisc.edu/wxwise/class/aos340/spr00/whatismatlab.htm
[accessed 24 November, 2016]

[L2] OpenCV (2016), About OpenCV, [online], available:
http://opencv.org/about.html
[accessed 24 November, 2016]

[L3] SimpleCV (2016), SimpleCV Tutorial, [online], available:
http://tutorial.simplecv.org/en/latest/
[accessed 24 November, 2016]

[L4] BeDjango (2017), Top 5 sites built with Django Framework, [online], available:
http://www.bedjango.com/blog/top-5-sites-built-django-framework/
[accessed 5 February, 2017]

[L5] Full Stack Python (2017), Flask, [online], available:
https://www.fullstackpython.com/flask.html
[accessed 5 February, 2017]