University of Huddersfield

MEng Group Project

Cryptic Crossword Solver

Problem Analysis Document

Authors:
Mohammad Rahman
Leanne Butcher
Stuart Leader
Luke Hackett

Supervisor:
Dr. Gary Allen
Examiner:
Dr. Sotirios Batsakis
Moderator:
Dr. Colin Venters

Friday, 9^{th} May 2014

Contents

List of Figures

Chapter 1

Problem Analysis

An initial problem analysis has been conducted to ensure that the overall project remains focused upon the original problem.

This chapter will discuss key topics and will recommend that these topics are researched further to help support the project. The problem analysis will also define the problem in more detail, to help with understanding the project and its purpose.

1.1 Problem

A typical crossword involves a grid with black squares, not to be filled in by the solver, and white squares used by the solver to input their answers. The input comes from the solver working out the answer to given clues. These clues have an orientation, a length and a number associated with its related position within the grid. The fundamental difference between a typical crossword and a cryptic crossword are the clues themselves.

Cryptic crosswords are a popular type of puzzles found in many parts of the world. Most common-wealth national newspapers will print cryptic crosswords of varying difficulty on a daily basis.

Cryptic crosswords are a unique style of crosswords, in which the answer to each given clue is a word puzzle. An answer can only be obtained if the cryptic clue is read in the correct way. Often when the clue is surface read, the clue makes no sense at all. The challenge is to find a way in which the reading of the clue leads to a solution. To aid with solving cryptic crosswords, the clues are written to be within specific categories, such as reversals and anagrams, which have individual characteristics.

Many users can often become frustrated when a clue appears to be unsolvable. It is the vast range of possible clues that often makes solving not only challenging but interesting as well.

Fundamentally, the overall aim of this project is to develop a piece of software that is able to solve any given type of cryptic crossword clue.

1.2 Product

Within this group project, three components will be delivered. The first deliverable is the final, working piece of software. Whilst the second and third deliverables are written reports. The second deliverable is a group written report comprising of the all research and implementation details of the software product. The final deliverable will be each member's individual analysis and evaluation of the project as a whole.

Based upon the given background and problem information it could be possible to develop a product that is able to solve the given problem.

The final product would be a piece of software that is able to understand a given clue and try to deduce what the answer to the clue is. This would require the software to have some form of natural language processing component as well as one or more cryptic crossword algorithms. Once a clue has been correctly "guessed" it can simply be returned to the user. It is the "guessing" of the answer that this project will primarily focus upon.

In order to gain maximum user coverage, the software must have an easy to use interface. The main reason for this is that the computer literacy of the intended users is not known - although basic computer literacy is assumed.

1.3 Client & Stakeholders

Dr Hugh Osborne, a lecturer from the University of Huddersfield will be the client for the group project. Dr Hugh Osborne has a keen interest in cryptic crosswords and the problem in the area which the group intends to help to eliminate. The role of the client for the group project will be to input ideas and potential requirements which Hugh, as an experienced solver of cryptic crosswords, would consider to be necessary. As a client for the project, Hugh will also be present for academic demonstrations.

Dr Gary Allen, Dr Sotirios Batsakis and Dr Colin Venters, all lecturers at the University of Huddersfield, will act as stakeholders for the group project. Gary will be the most involved external individual as the project supervisor. The role of the project supervisor requires frequent meetings with the team to monitor the development of the project, provide guidance as well as opinion on certain aspects of the life cycle.

Sotirios and Colin will have a less active role within the project during the project life cycle than the role of the client or the project supervisor. These stakeholders will play active roles at particular milestones of the group project such as providing guidance for the proposal

of the project and at the project demonstration approximately half way through the life cycle.

1.4 Users

Kathryn Friedlander and Philip Fine (Friedlander and Fine, 2009) carried out an investigation into whether the amount of cryptic crosswords completed by a solver determined how successful they were at solving them. To complete this study they gathered data from 241 people and have deduced the following facts about the user base (Friedlander and Fine, 2009):

- "209 M, 32 F"
- "mean age=53 years, range=23 – 83"
- "mean time spent=8 hours per week, range=1-30"

To support decisions made within the project life cycle an additional quantitative research method has been utilized to gain a larger understanding of the types of users the deliverable will attract. The research method used by the team is in the form of a survey.

The survey results gathered were seen as additional justifications for the purpose of the project. Moreover, data collected from the survey was expected to indicate the locations in which users complete cryptic crosswords to understand the potential need for the deliverable to be of a transportable nature.

The following questions were asked:

1. Do you play cryptic crosswords?
2. How often do you play?
3. How do you complete cryptic crosswords?
4. Do you often finish them?
5. If no to the previous question, what reason don't you finish them?
6. What is your age group?
7. When do you play cryptic crosswords?
8. What gender are you?
9. What is the highest qualification you have?
10. What platform is your mobile phone on?

The survey was conducted between 18th November and 4th January. It was distributed across the University of Huddersfield portal message board, Facebook and Twitter.

1.4.1 Results

The survey that was outlined within the previous section managed to provide some interesting trends. Within this subsection a number of the trends will be highlighted, and discussed. It is hoped that the trends will be able to help guide the research, design and development processes throughout the project.

One of the key questions that was asked within the survey was "how do you complete cryptic crosswords?", which the question referred to the format in which users generally tended to play or favour.

Figure 1.1 illustrates the responses from the question. The overall statement is that 56% of those surveyed favoured a paper based format which perhaps is a little surprising in today's 'digital age'.

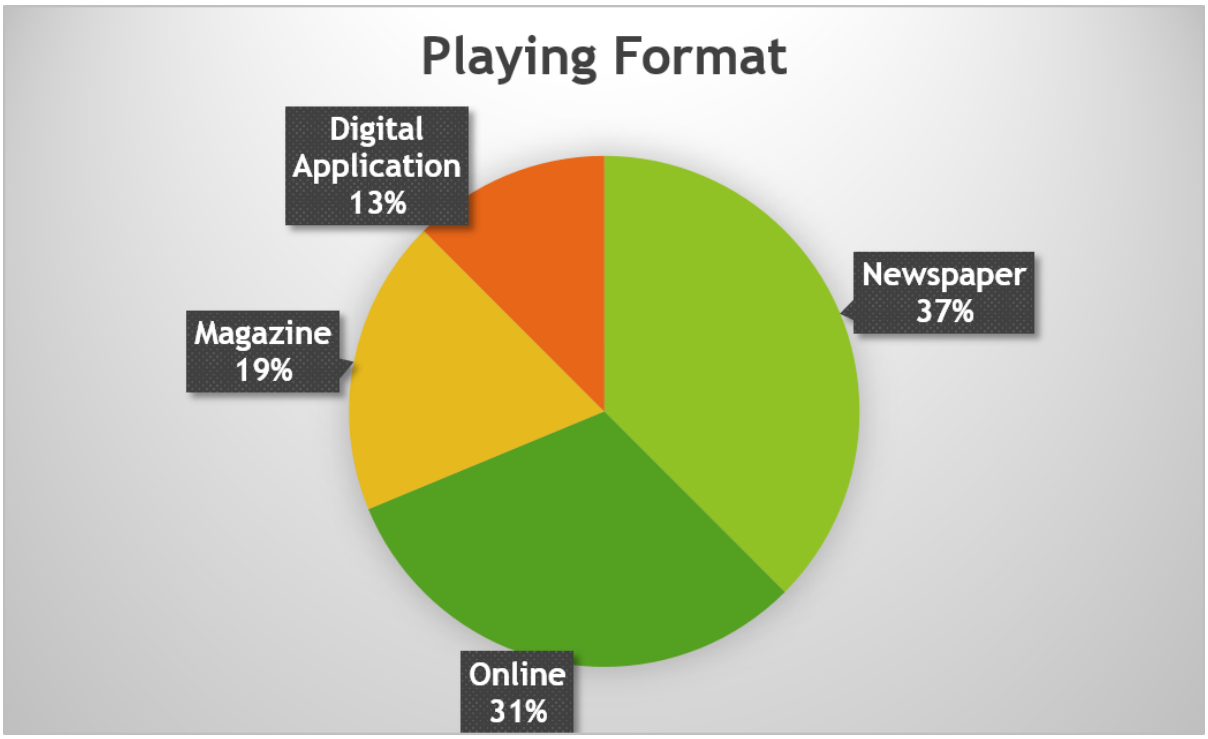

Figure 1.1: Pie chart illustrating how users complete cryptic crosswords

Following on from this it was also surprising that very few individuals use a mobile-based application to complete cryptic crosswords — especially as on-line crosswords were the second

most used format. It is this area that is intended to be a large basis for this project, and hence this will need to be investigated thoroughly.

The survey also tried to deduce the reasons behind why cryptic crossword users are perhaps sometimes unable to solve a clue. The understanding of the responses to this question is a critical part of the project, as that it is intended that the final product should be able to solve a given clue, and thus help a solver.

Figure 1.2 illustrates the responses from all users — both advanced and basic users. The chart shows that the most common reason as to why users are unable to complete crosswords is that they don't have the time to complete the crossword — something that the project may not be able to directly solve.

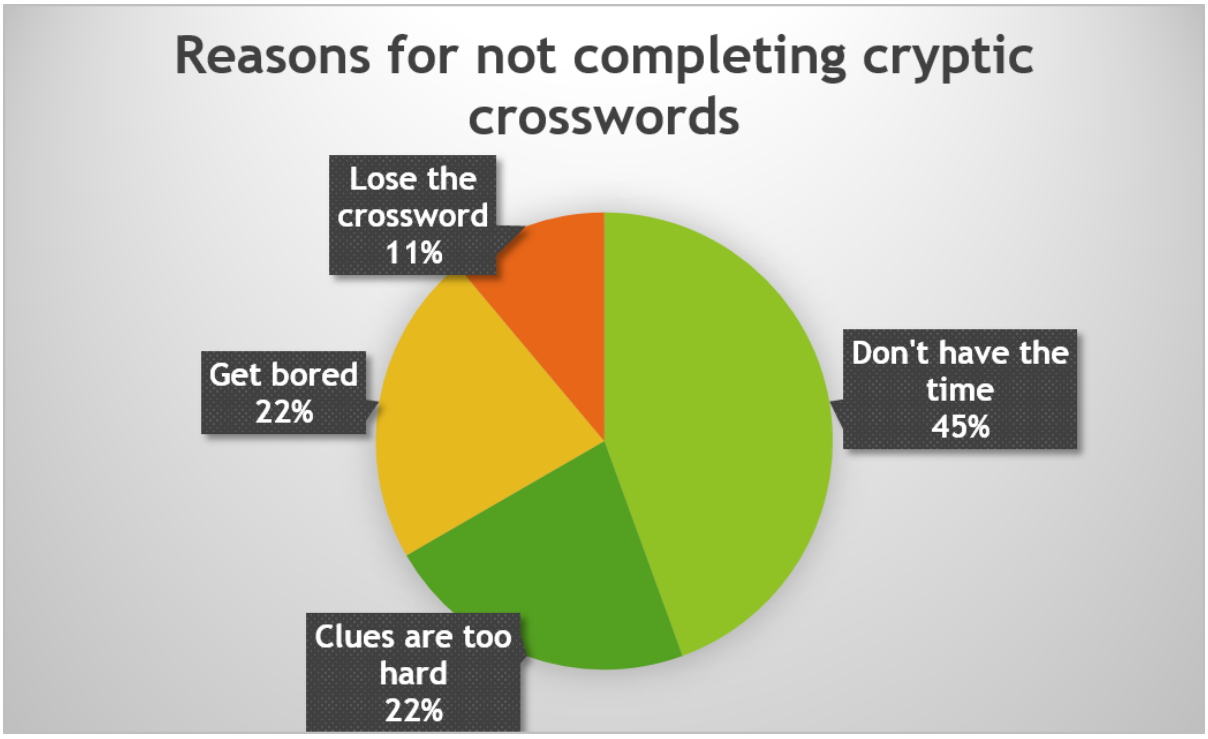

Figure 1.2: Pie chart illustrating why all users are unable to complete cryptic crosswords

However the next top two responses — "clues are too hard" and "get bored" — could be attempted to be solved within this project. These two responses could be linked together, for example do users get bored because the clues are too hard? Although the survey did not highlight this, it did illustrate a potential trend in the data that can be investigated further.

Figure 1.2 illustrated responses from all users, if "advanced users" are removed then a more clear trend may emerge. For the purposes of this test an "advanced user" can be defined as a user who regularly tries to solve a cryptic crossword.

Figure 1.3 illustrates responses only from "basic users". Although the overall trend is the same, it does highlight a larger gap between the two other responses — "clues are too hard" and "get bored".

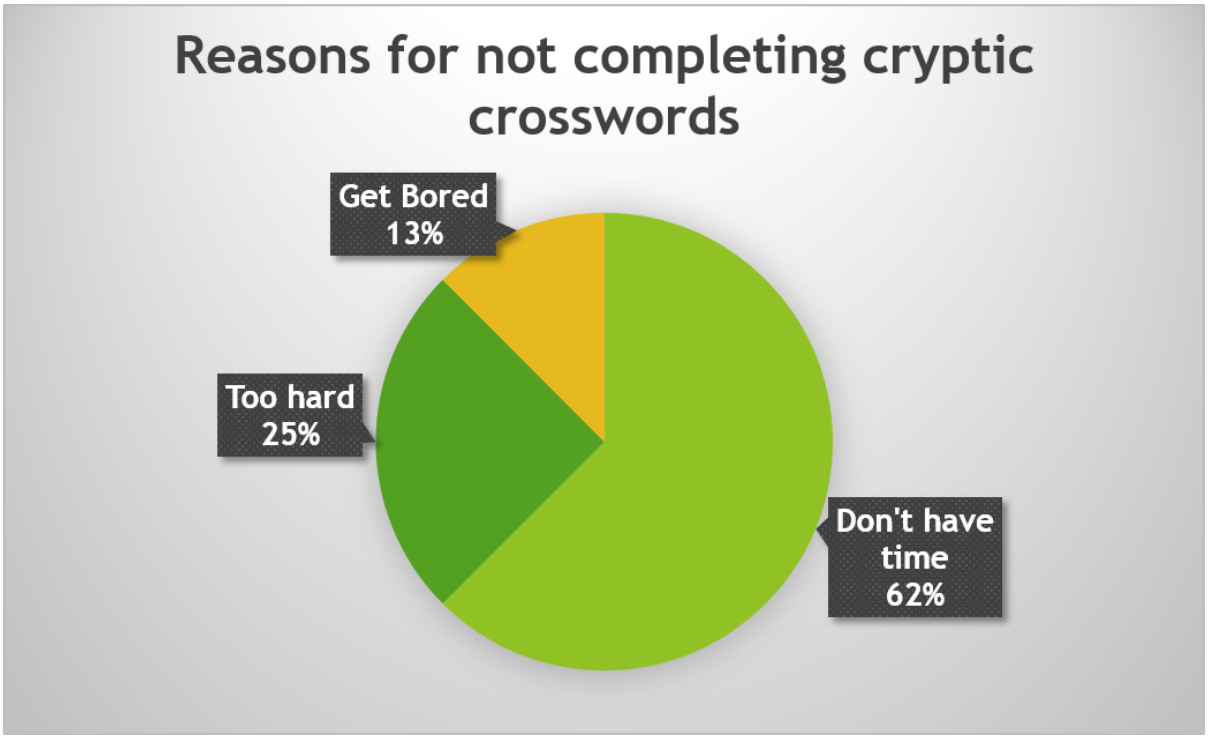

Figure 1.3: Pie chart illustrating why basic users are unable to complete cryptic crosswords

This identifies another possible trend that users who can't solve clues may wish to 'learn' how to solve that clue. This factor will need to be taken into consideration through the remainder of the project.

Additionally the survey tried to deduce when cryptic crosswords are completed. Figure 1.4 illustrates the responses. Generally speaking it could be said that cryptic crosswords are completed to "pass time". The top two answers indicated that users completed crosswords when "bored" or "travelling".

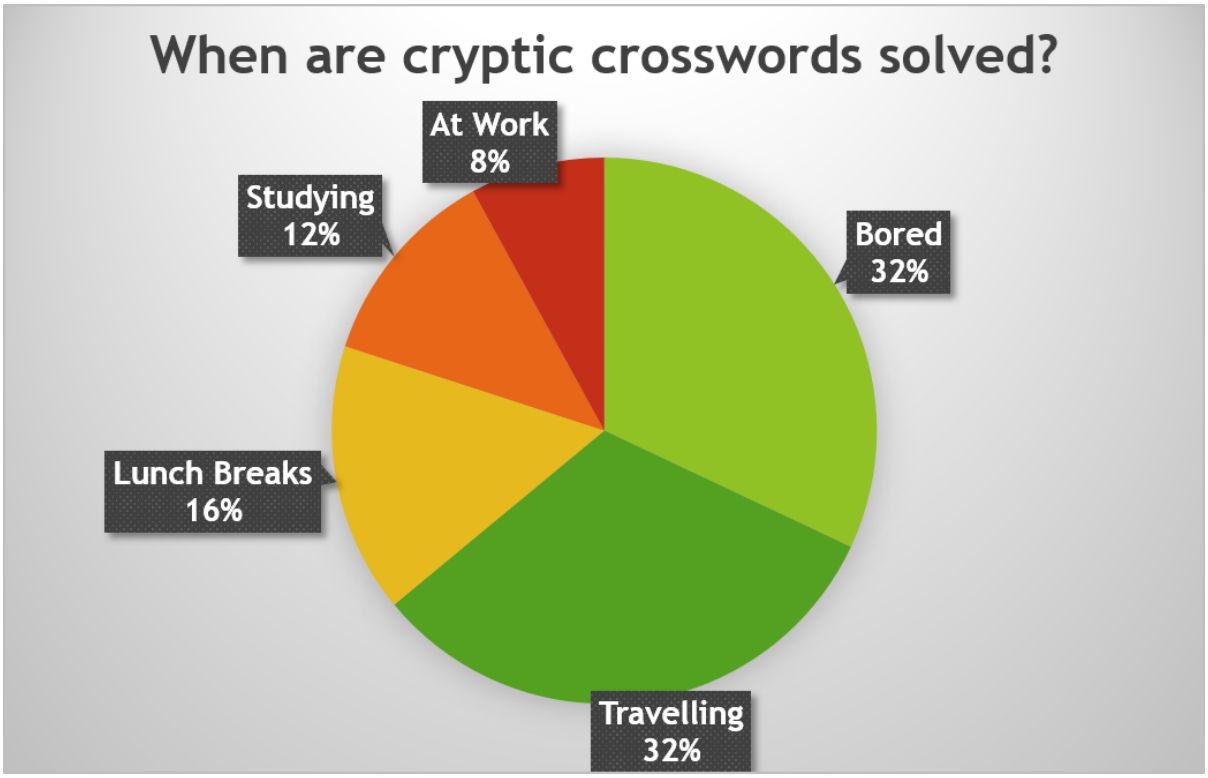

Figure 1.4: Pie chart illustrating when cryptic crosswords are solved

The responses to the question were quite important, as it highlights that the responses are people who could be categorised as people who are "on the go". This means that they are not fixed in one location, such as travelling from home to work and back home again. Furthermore it highlights that there may be a potential gap in the market for a mobile based application to aid in the solving of cryptic crosswords.

With this in mind, the final question tried to deduce which mobile platforms users owned and used. Figure 1.5 illustrates the responses, and unsurprisingly the top two platform choices was Apple's iOS and Google's Android operating systems.

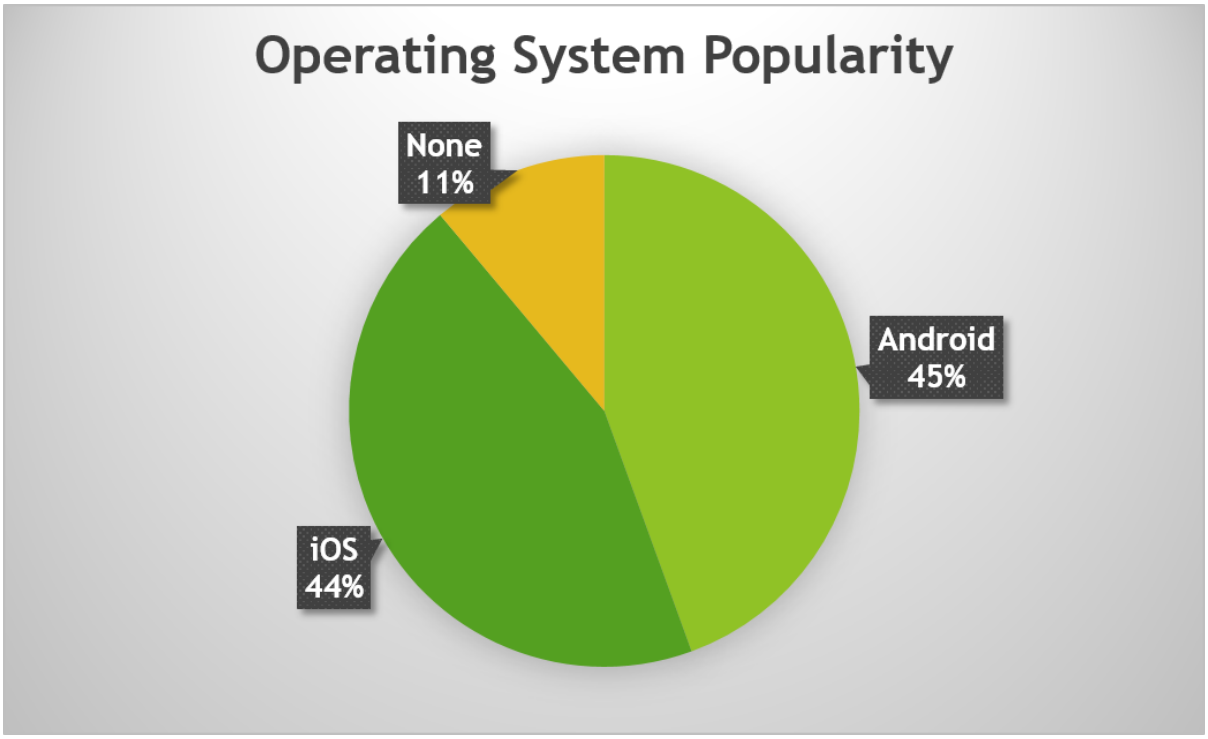

Figure 1.5: Pie chart illustrating the ownership of mobile devices by operating system

The chart indicates that if a mobile-based application was feasible then the application should run on at least iOS or Android.

The responses and results presented from the survey will be used to direct the 'formal' academic research that will form part of the written deliverable of the project. The specific research areas will be covered in the following section.

1.5 Research Areas

Before undertaking the project an initial review was conducted. The review's objective was to determine the feasibility of the project as a whole. The review also covered whether or not the project has been completed before.

From the outlined background and problem information it is clear that cryptic crosswords are a popular form of entertainment. It is also clear that some clues are particularly difficult to solve, and users may often ask other people for help in solving a given clue.

1.5.1 Cryptic Crosswords

A review of the national UK newspapers was conducted to determine whether or not there is a pattern in cryptic crosswords. Of all the newspaper's websites that were reviewed (The Guardian, The Times, The Independent and The Mirror) it was clear that all cryptic crosswords are of the same style.

Each clue is categorised as being either 'across' or 'down' with its corresponding grid number, as well as containing the number of letters the answer should be. An example is shown below:

12. The seamstress's sensation? (4, 3, 7) =>PINS AND NEEDLES

The Guardian's website utilises web standard technologies such as HTML and CSS, and also provides an option to solve a clue. The Mirror's website follows a similar approach to the Guardian's website; however solutions can only be obtained by dialing a premium telephone number.

The Times and the Independent both utilise a different approach and that is to serve a Java applet. Both Java applets allow the user to solve a clue should they get stuck. The Times provides puzzles as part of their paid subscription service.

All of the above newspapers publish cryptic crosswords upon a daily basis, with the solutions to the crosswords appearing in the next day's newspaper.

Following from the crossword review, a second review into cryptic crossword solvers was undertaken. The objective of this review was to determine whether or not computerised cryptic crossword solvers exist. The three cryptic crossword solvers that were identified were One Across, Crossword Tools and Cryptic Solver.

Each of the solvers manages to solve some clues with the same answers, with other clues providing a range of possible answers.

Crossword Tools (Crossword Tools, 2013) is a paid subscription based service, which allows users to enter a clue and a pattern. A pattern can contain part of the answer or the number

of letters the answer has. If multiple answers are available, they are displayed. An example is shown below:

Kind of dog (10) =>the answer is 10 letters long.

Kind of dog (?????????r) =>the answer is 10 letters long, final letter is 'r'.

Cryptic Solver (Cryptic Solver, 2013) is a free service that offers the same functionality as Crossword Tools. Although Cryptic Solver does provide the correct answer, it does not necessarily provide the correct answer at the top of the list.

Finally One Across (One Across, 2013) provides all the same functionality as the previous two solvers, along with a score. The score is linked to the number of people who have used the given answer (effectively it's a ratings system). One Across uniquely highlights how it has managed to deduce the answer, showing the break downs of each sentence. As with Cryptic Solver, One Across is a free service that doesn't require a subscription.

1.5.2 Natural Language Processing

In order to correctly solve a clue, some form of natural language processing may be required. It is the natural language processing that could try to deduce the meaning of a clue. It is the meaning that can then be aligned with possible answers.

An example of natural language processing can be found within the One Across application. Given a clue (and a pattern) it will try to provide an accurate solution:

Spin broken shingle (7) =>ENGLISH

In order for the answer to be obtained, One Across will follow a natural language processing path and will provide it's trace path. The trace path shows how the clue has been broken down to get to the answer. The trace path for the above clue can be found below:

'spin' is the definition.

'broken' means to anagram 'shingle' to get ENGLISH.

ENGLISH matches 'spin' with confidence score 100%.

1.5.3 Application Platform

The existing products that have been discussed within this problem analysis have all been accessible via a browser. Although this is an acceptable platform, there could be a better platform that allows users to utilise the technology easier.

As previously mentioned, most crosswords are designed for users who have a few minutes to spare on the move. With the recent trends in owning a smartphone or tablet, there may be a gap in the market for a high quality mobile cryptic crossword solver.

Additionally it may also be the case that the system will need to be distributed across several machines. The main reason for this could be the fact that the system is too complex for it to be able to run efficiently upon a mobile device. It is with this in mind that a review in to the various system architectures will need to be conducted along side the application format.

An in-depth review of all possible approaches will need to be conducted in order to deduce the viability of this proposal.

Glossary of Terms

The following section contains a glossary with the meanings of all names, acronyms, and abbreviations used by the stakeholders.

Term/Acronym	**Definition**
The Guardian	A national UK newspaper that prints daily cryptic crosswords
Android	A mobile phone software platform by Google Inc.
BlackBerry	A mobile phone hardware and software platform developed by BlackBerry Limited
iOS	A mobile phone software platform developed by Apple Inc.
iPhone	A smart phone developed by Apple Inc.
iPod	A portable digital music player developed by Apple Inc.
iPad	A tablet developed by Apple Inc.
NLP	Natural Language Processing
SRS	Software Requirements Specification
App	Shorthand for application

Bibliography

Crossword Tools (2013). Clue solver. Retrieved from `http://www.crosswordtools.com/cm/`.

Cryptic Solver (2013). Solver. Retrieved from `http://cryptic-solver.appspot.com/`.

Friedlander, K. and Fine, P. (2009). Expertise in cryptic crossword performance. In *An exploratory survey*, Auckland, New Zealand. International Symposium on Performance Science.

One Across (2013). Crossword puzzle help. Retrieved from `http://www.oneacross.com/`.