Karina Palyutina

Machine learning inference of search engine heuristics

Part II Project

St Catharine's College

February 18, 2013

Proforma

Name:	**Karina Palyutina**
College:	**St Catharine's College**
Project Title:	**Machine learning inference of search engine heuris**
Examination:	**Part II Project**
Word Count:	[1] **(well less than the 12000 limit)**
Project Originator:	Dr Jon Crowcroft
Supervisor:	Dr Jon Crowcroft

Original Aims of the Project

To write a demonstration dissertation[2] using LaTeX to save student's time when writing their own dissertations. The dissertation should illustrate how to use the more common LaTeX constructs. It should include pictures and diagrams to show how these can be incorporated into the dissertation. It should contain the entire LaTeX source of the dissertation and the Makefile. It should explain how to construct an MSDOS disk of the dissertation in Postscript format that can be used by the book shop for printing, and, finally, it should have the prescribed layout and format of a diploma dissertation.

Work Completed

All that has been completed appears in this dissertation.

[1] This word count was computed by `detex diss.tex | tr -cd '0-9A-Za-z \n' | wc -w`

[2] A normal footnote without the complication of being in a table.

Special Difficulties

Learning how to incorporate encapulated postscript into a LaTeX document on both CUS and Thor.

Declaration

I, [Name] of [College], being a candidate for Part II of the Computer Science Tripos [or the Diploma in Computer Science], hereby declare that this dissertation and the work described in it are my own work, unaided except as may be specified below, and that the dissertation does not contain material that has already been used to any substantial extent for a comparable purpose.

Signed [signature]

Date [date]

Contents

List of Figures

Acknowledgements

This document owes much to an earlier version written by Simon Moore [2]. His help, encouragement and advice was greatly appreciated.

Chapter 1

Introduction

1.1 Overview of the files

This document consists of the following files:

- `Makefile` — The Makefile for the dissertation and Project Proposal
- `diss.tex` — The dissertation
- `propbody.tex` — Appendix C – the project proposal
- `proposal.tex` — A LaTeX main file for the proposal
- `figs` – A directory containing diagrams and pictures
- `refs.bib` — The bibliography database

1.2 Building the document

This document was produced using LaTeX 2$_{\varepsilon}$which is based upon LaTeX[1]. To build the document you first need to generate `diss.aux` which, amongst other things, contains the references used. This if done by executing the command:

```
latex diss
```

Then the bibliography can be generated from `refs.bib` using:

```
bibtex diss
```

Finally, to ensure all the page numbering is correct run `latex` on `diss.tex` until the `.aux` files do not change. This usually takes 2 more runs.

1.2.1 The makefile

To simplify the calls to `latex` and `bibtex`, a makefile has been provided, see Appendix B.1. It provides the following facilities:

- `make`
  Display help information.

- `make prop`
  Run `latex proposal; xdvi proposal.dvi`.

- `make diss.ps`
  Make the file `diss.ps`.

- `make gv`
  View the dissertation using ghostview after performing `make diss.ps`, if necessary.

- `make gs`
  View the dissertation using ghostscript after performing `make diss.ps`, if necessary.

- `make count`
  Display an estimate of the word count.

- `make all`
  Construct `proposal.dvi` and `diss.ps`.

- `make pub`
  Make a `.tar` version of the `demodiss` directory and place it in my `public_html` directory.

- `make clean`
  Delete all files except the source files of the dissertation. All these deleted files can be reconstructed by typing `make all`.

- `make pr`
  Print the dissertation on your default printer.

1.3 Counting words

An approximate word count of the body of the dissertation may be obtained using:

```
   wc diss.tex
```

Alternatively, try something like:

```
   detex diss.tex | tr -cd '0-9A-Z a-z\n' | wc -w
```

Chapter 2

Preparation

This chapter is empty!

Chapter 3

Implementation

3.1 Verbatim text

Verbatim text can be included using \begin{verbatim} and \end{verbatim}. I normally use a slightly smaller font and often squeeze the lines a little closer together, as in:

```
GET "libhdr"

GLOBAL { count:200; all  }

LET try(ld, row, rd) BE TEST row=all
                        THEN count := count + 1
                        ELSE { LET poss = all & ~(ld | row | rd)
                               UNTIL poss=0 DO
                               { LET p = poss & -poss
                                 poss := poss - p
                                 try(ld+p << 1, row+p, rd+p >> 1)
                               }
                             }
LET start() = VALOF
{ all := 1
  FOR i = 1 TO 12 DO
  { count := 0
    try(0, 0, 0)
    writef("Number of solutions to %i2-queens is %i5*n", i, count)
    all := 2*all + 1
  }
  RESULTIS 0
}
```

3.2 Tables

Here is a simple example[1] of a table.

Left	Centred	Right
Justified		Justified
First	A	XXX
Second	AA	XX
Last	AAA	X

There is another example table in the proforma.

3.3 Simple diagrams

Simple diagrams can be written directly in LaTeX. For example, see figure 3.1 on page 9 and see figure 3.2 on page 9.

3.4 Adding more complicated graphics

The use of LaTeX format can be tedious and it is often better to use encapsulated postscript to represent complicated graphics. Figure 3.3 and 3.5 on page 11 are examples. The second figure was drawn using `xfig` and exported in `.eps` format. This is my recommended way of drawing all diagrams.

[1] A footnote

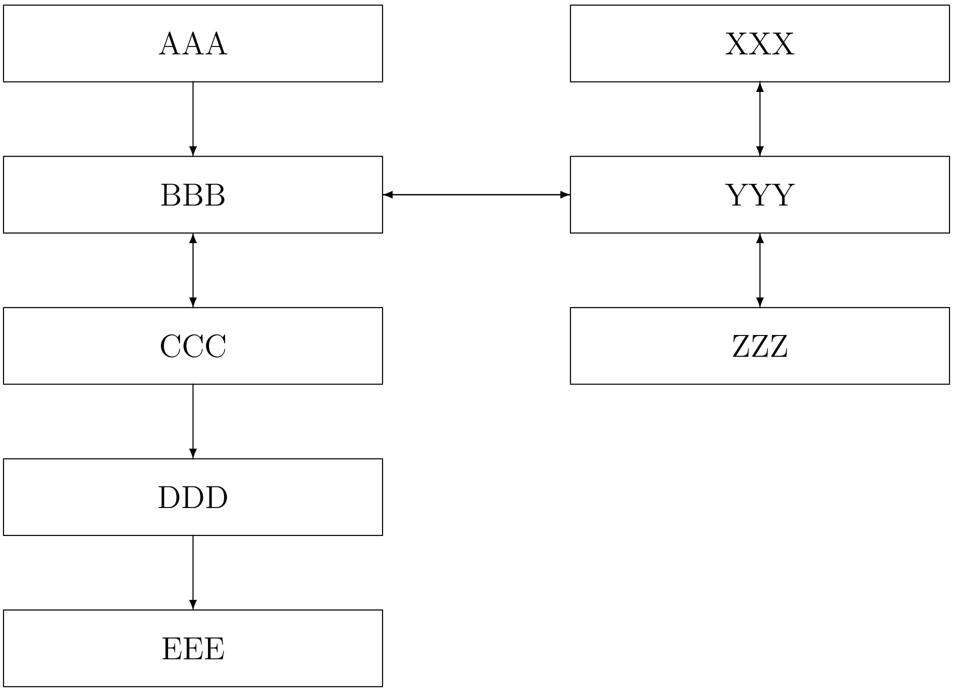

Figure 3.1: A picture composed of boxes and vectors.

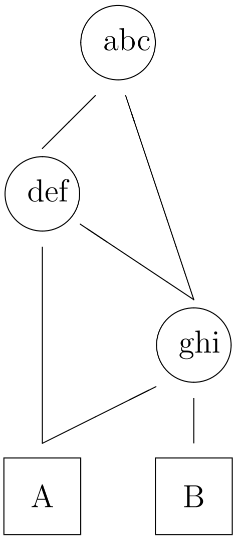

Figure 3.2: A diagram composed of circles, lines and boxes.

Figure 3.3: Example figure using encapsulated postscript

Figure 3.4: Example figure where a picture can be pasted in

Figure 3.5: Example diagram drawn using `xfig`

Chapter 4

Evaluation

4.1 Printing and binding

If you have access to a laser printer that can print on two sides, you can use it to print two copies of your dissertation and then get them bound by the Computer Laboratory Bookshop. Otherwise, print your dissertation single sided and get the Bookshop to copy and bind it double sided.

Better printing quality can sometimes be obtained by giving the Bookshop an MSDOS 1.44 Mbyte 3.5" floppy disc containing the Postscript form of your dissertation. If the file is too large a compressed version with `zip` but not `gnuzip` nor `compress` is acceptable. However they prefer the uncompressed form if possible. From my experience I do not recommend this method.

4.1.1 Things to note

- Ensure that there are the correct number of blank pages inserted so that each double sided page has a front and a back. So, for example, the title page must be followed by an absolutely blank page (not even a page number).

- Submitted postscript introduces more potential problems. Therefore you must either allow two iterations of the binding process (once in a digital form, falling back to a second, paper, submission if necessary) or submit both paper and electronic versions.

- There may be unexpected problems with fonts.

4.2 Further information

See the Computer Lab's world wide web pages at URL:
`http://www.cl.cam.ac.uk/TeXdoc/TeXdocs.html`

Chapter 5

Conclusion

I hope that this rough guide to writing a dissertation is LaTeX has been helpful and saved you time.

Bibliography

[1] L. Lamport. *LaTeX — a document preparation system — user's guide and reference manual*. Addison-Wesley, 1986.

[2] S.W. Moore. How to prepare a dissertation in latex, 1995.

Appendix A

Latex source

A.1 diss.tex

```
% The master copy of this demo dissertation is held on my filespace
% on the cl file serve (/homes/mr/teaching/demodissert/)

% Last updated by MR on 2 August 2001

\documentclass[12pt,twoside,notitlepage]{report}

\usepackage{a4}
\usepackage{verbatim}

\input{epsf}                            % to allow postscript inclusions
% On thor and CUS read top of file:
%     /opt/TeX/lib/texmf/tex/dvips/epsf.sty
% On CL machines read:
%     /usr/lib/tex/macros/dvips/epsf.tex



\raggedbottom                           % try to avoid widows and orphans
\sloppy
\clubpenalty1000%
\widowpenalty1000%

\addtolength{\oddsidemargin}{6mm}       % adjust margins
\addtolength{\evensidemargin}{-8mm}

\renewcommand{\baselinestretch}{1.1}    % adjust line spacing to make
                                        % more readable

\begin{document}

\bibliographystyle{plain}


%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Title
```

```
\pagestyle{empty}

\hfill{\LARGE \bf Karina Palyutina}

\vspace*{60mm}
\begin{center}
\Huge
{\bf Machine learning inference of search engine heuristics} \\
\vspace*{5mm}
Part II Project \\
\vspace*{5mm}
St Catharine’s College \\
\vspace*{5mm}
\today  % today’s date
\end{center}

\cleardoublepage

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Proforma, table of contents and list of figures

\setcounter{page}{1}
\pagenumbering{roman}
\pagestyle{plain}

\chapter*{Proforma}

{\large
\begin{tabular}{ll}
Name:               & \bf Karina Palyutina                       \\
College:            & \bf St Catharine’s College                     \\
Project Title:      & \bf Machine learning inference of search engine heuristics \\
Examination:        & \bf Part II Project        \\
Word Count:         & \bf \footnotemark[1]
(well less than the 12000 limit) \\
Project Originator: & Dr Jon Crowcroft                    \\
Supervisor:         & Dr Jon Crowcroft                  \\
\end{tabular}
}
\footnotetext[1]{This word count was computed
by {\tt detex diss.tex | tr -cd ’0-9A-Za-z $\tt\backslash$n’ | wc -w}
}
\stepcounter{footnote}


\section*{Original Aims of the Project}

To write a demonstration dissertation\footnote{A normal footnote without the
complication of being in a table.} using \LaTeX\ to save
student’s time when writing their own dissertations. The dissertation
should illustrate how to use the more common \LaTeX\ constructs. It
should include pictures and diagrams to show how these can be
incorporated into the dissertation.  It should contain the entire
\LaTeX\ source of the dissertation and the Makefile.  It should
explain how to construct an MSDOS disk of the dissertation in
Postscript format that can be used by the book shop for printing, and,
finally, it should have the prescribed layout and format of a diploma
dissertation.
```

```
\section*{Work Completed}

All that has been completed appears in this dissertation.

\section*{Special Difficulties}

Learning how to incorporate encapulated postscript into a \LaTeX\
document on both CUS and Thor.

\newpage
\section*{Declaration}

I, [Name] of [College], being a candidate for Part II of the Computer
Science Tripos [or the Diploma in Computer Science], hereby declare
that this dissertation and the work described in it are my own work,
unaided except as may be specified below, and that the dissertation
does not contain material that has already been used to any substantial
extent for a comparable purpose.

\bigskip
\leftline{Signed [signature]}

\medskip
\leftline{Date [date]}

\cleardoublepage

\tableofcontents

\listoffigures

\newpage
\section*{Acknowledgements}

This document owes much to an earlier version written by Simon Moore
\cite{moore95}.  His help, encouragement and advice was greatly
appreciated.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% now for the chapters

\cleardoublepage        % just to make sure before the page numbering
                        % is changed

\setcounter{page}{1}
\pagenumbering{arabic}
\pagestyle{headings}

\chapter{Introduction}

\section{Overview of the files}

This document consists of the following files:

\begin{itemize}
\item {\tt Makefile} --- The Makefile for the dissertation and Project Proposal
\item {\tt diss.tex} --- The dissertation
```

```
\item {\tt propbody.tex} --- Appendix~C  -- the project proposal
\item {\tt proposal.tex}  --- A \LaTeX\ main file for the proposal
\item{\tt figs} -- A directory containing diagrams and pictures
\item{\tt refs.bib} --- The bibliography database
\end{itemize}

\section{Building the document}

This document was produced using \LaTeXe which is based upon
\LaTeX\cite{Lamport86}.  To build the document you first need to
generate {\tt diss.aux} which, amongst other things, contains the
references used.  This if done by executing the command:

{\tt latex diss}

\noindent
Then the bibliography can be generated from {\tt refs.bib} using:

{\tt bibtex diss}

\noindent
Finally, to ensure all the page numbering is correct run {\tt latex}
on {\tt diss.tex} until the {\tt .aux} files do not change.  This
usually takes 2 more runs.

\subsection{The makefile}

To simplify the calls to {\tt latex} and {\tt bibtex},
a makefile has been provided, see Appendix~\ref{makefile}.
It provides the following facilities:

\begin{itemize}

\item{\tt make} \\
 Display help information.

\item{\tt make prop} \\
 Run {\tt latex proposal; xdvi proposal.dvi}.

\item{\tt make diss.ps} \\
 Make the file {\tt diss.ps}.

\item{\tt make gv} \\
 View the dissertation using ghostview after performing
{\tt make diss.ps}, if necessary.

\item{\tt make gs} \\
 View the dissertation using ghostscript after performing
{\tt make diss.ps}, if necessary.

\item{\tt make count} \\
Display an estimate of the word count.

\item{\tt make all} \\
Construct {\tt proposal.dvi} and {\tt diss.ps}.

\item{\tt make pub} \\ Make a {\tt .tar} version of the {\tt demodiss}
directory and place it in my {\tt public\_html} directory.
```

```
\item{\tt make clean} \\ Delete all files except the source files of
the dissertation. All these deleted files can be reconstructed by
typing {\tt make all}.

\item{\tt make pr} \\
Print the dissertation on your default printer.

\end{itemize}


\section{Counting words}

An approximate word count of the body of the dissertation may be
obtained using:

{\tt wc diss.tex}

\noindent
Alternatively, try something like:

\verb/detex diss.tex | tr -cd ’0-9A-Z a-z\n’ | wc -w/




\cleardoublepage



\chapter{Preparation}

This chapter is empty!


\cleardoublepage
\chapter{Implementation}

\section{Verbatim text}

Verbatim text can be included using \verb|\begin{verbatim}| and
\verb|\end{verbatim}|. I normally use a slightly smaller font and
often squeeze the lines a little closer together, as in:

{\renewcommand{\baselinestretch}{0.8}\small\begin{verbatim}
GET "libhdr"

GLOBAL { count:200; all  }

LET try(ld, row, rd) BE TEST row=all
                        THEN count := count + 1
                        ELSE { LET poss = all & ~(ld | row | rd)
                               UNTIL poss=0 DO
                               { LET p = poss & -poss
                                 poss := poss - p
                                 try(ld+p << 1, row+p, rd+p >> 1)
                               }
                             }
LET start() = VALOF
{ all := 1
```

```
  FOR i = 1 TO 12 DO
  { count := 0
    try(0, 0, 0)
    writef("Number of solutions to %i2-queens is %i5*n", i, count)
    all := 2*all + 1
  }
  RESULTIS 0
}
\end{verbatim}
}

\section{Tables}

\begin{samepage}
Here is a simple example\footnote{A footnote} of a table.

\begin{center}
\begin{tabular}{l|c|r}
Left      & Centred & Right \\
Justified &         & Justified \\[3mm]
%\hline\\%[-2mm]
First     & A       & XXX \\
Second    & AA      & XX  \\
Last      & AAA     & X   \\
\end{tabular}
\end{center}

\noindent
There is another example table in the proforma.
\end{samepage}

\section{Simple diagrams}

Simple diagrams can be written directly in \LaTeX.  For example, see
figure~\ref{latexpic1} on page~\pageref{latexpic1} and see
figure~\ref{latexpic2} on page~\pageref{latexpic2}.

\begin{figure}
\setlength{\unitlength}{1mm}
\begin{center}
\begin{picture}(125,100)
\put(0,80){\framebox(50,10){AAA}}
\put(0,60){\framebox(50,10){BBB}}
\put(0,40){\framebox(50,10){CCC}}
\put(0,20){\framebox(50,10){DDD}}
\put(0,00){\framebox(50,10){EEE}}

\put(75,80){\framebox(50,10){XXX}}
\put(75,60){\framebox(50,10){YYY}}
\put(75,40){\framebox(50,10){ZZZ}}

\put(25,80){\vector(0,-1){10}}
\put(25,60){\vector(0,-1){10}}
\put(25,50){\vector(0,1){10}}
\put(25,40){\vector(0,-1){10}}
\put(25,20){\vector(0,-1){10}}

\put(100,80){\vector(0,-1){10}}
\put(100,70){\vector(0,1){10}}
```

```
\put(100,60){\vector(0,-1){10}}
\put(100,50){\vector(0,1){10}}

\put(50,65){\vector(1,0){25}}
\put(75,65){\vector(-1,0){25}}
\end{picture}
\end{center}
\caption{\label{latexpic1}A picture composed of boxes and vectors.}
\end{figure}

\begin{figure}
\setlength{\unitlength}{1mm}
\begin{center}

\begin{picture}(100,70)
\put(47,65){\circle{10}}
\put(45,64){abc}

\put(37,45){\circle{10}}
\put(37,51){\line(1,1){7}}
\put(35,44){def}

\put(57,25){\circle{10}}
\put(57,31){\line(-1,3){9}}
\put(57,31){\line(-3,2){15}}
\put(55,24){ghi}

\put(32,0){\framebox(10,10){A}}
\put(52,0){\framebox(10,10){B}}
\put(37,12){\line(0,1){26}}
\put(37,12){\line(2,1){15}}
\put(57,12){\line(0,2){6}}
\end{picture}

\end{center}
\caption{\label{latexpic2}A diagram composed of circles, lines and boxes.}
\end{figure}



\section{Adding more complicated graphics}

The use of \LaTeX\ format can be tedious and it is often better to use
encapsulated postscript to represent complicated graphics.
Figure~\ref{epsfig} and ~\ref{xfig} on page \pageref{xfig} are
examples. The second figure was drawn using {\tt xfig} and exported in
{\tt.eps} format. This is my recommended way of drawing all diagrams.


\begin{figure}[tbh]
\centerline{\epsfbox{figs/cuarms.eps}}
\caption{\label{epsfig}Example figure using encapsulated postscript}
\end{figure}

\begin{figure}[tbh]
\vspace{4in}
\caption{\label{pastedfig}Example figure where a picture can be pasted in}
\end{figure}
```

```
\begin{figure}[tbh]
\centerline{\epsfbox{figs/diagram.eps}}
\caption{\label{xfig}Example diagram drawn using {\tt xfig}}
\end{figure}




\cleardoublepage
\chapter{Evaluation}

\section{Printing and binding}

If you have access to a laser printer that can print on two sides, you
can use it to print two copies of your dissertation and then get them
bound by the Computer Laboratory Bookshop. Otherwise, print your
dissertation single sided and get the Bookshop to copy and bind it double
sided.


Better printing quality can sometimes be obtained by giving the
Bookshop an MSDOS 1.44~Mbyte 3.5" floppy disc containing the
Postscript form of your dissertation. If the file is too large a
compressed version with {\tt zip} but not {\tt gnuzip} nor {\tt
compress} is acceptable. However they prefer the uncompressed form if
possible. From my experience I do not recommend this method.

\subsection{Things to note}

\begin{itemize}
\item Ensure that there are the correct number of blank pages inserted
so that each double sided page has a front and a back.  So, for
example, the title page must be followed by an absolutely blank page
(not even a page number).

\item Submitted postscript introduces more potential problems.
Therefore you must either allow two iterations of the binding process
(once in a digital form, falling back to a second, paper, submission if
necessary) or submit both paper and electronic versions.

\item There may be unexpected problems with fonts.

\end{itemize}

\section{Further information}

See the Computer Lab's world wide web pages at URL:

{\tt http://www.cl.cam.ac.uk/TeXdoc/TeXdocs.html}


\cleardoublepage
\chapter{Conclusion}

I hope that this rough guide to writing a dissertation is \LaTeX\ has
been helpful and saved you time.
```

```
\cleardoublepage

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% the bibliography

\addcontentsline{toc}{chapter}{Bibliography}
\bibliography{refs}
\cleardoublepage

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% the appendices
\appendix

\chapter{Latex source}

\section{diss.tex}
{\scriptsize\verbatiminput{diss.tex}}

\section{proposal.tex}
{\scriptsize\verbatiminput{proposal.tex}}

\section{propbody.tex}
{\scriptsize\verbatiminput{propbody.tex}}



\cleardoublepage

\chapter{Makefile}

\section{\label{makefile}Makefile}
{\scriptsize\verbatiminput{makefile.txt}}

\section{refs.bib}
{\scriptsize\verbatiminput{refs.bib}}


\cleardoublepage

\chapter{Project Proposal}

%\input{propbody}

\end{document}
```

A.2 proposal.tex

```
% This is a LaTeX driving document to produce a standalone copy
% of the project proposal held in propbody.tex.  Notice that
% propbody can be used in this context as well as being incorporated
% in the dissertation (see diss.tex).

\documentclass[12pt,a4,parskip=full]{scrartcl}
\usepackage{graphicx}
\usepackage[font={small,it}]{caption}
```

```
\usepackage{pdfpages}
\begin{document}
\includepdf[pages={-}]{cover.pdf}
\include{propbody}

\end{document}
```

A.3 propbody.tex

```
\title{Machine learning inference of search engine heuristics}
%\subtitle{Part II Computer Science Project Proposal}
\author{K. Palyutina, St. Catharine’s College \\
        Originator: Dr. Jon Crowcroft}

\maketitle


\vfil


\noindent
{\bf Project Supervisor:} Prof. J. Crowcroft
\vspace{0.2in}

\noindent
{\bf Director of Studies:} Dr. S. Taraskin
\vspace{0.2in}
\noindent

\noindent
{\bf Project Overseers:} Dr.~M.~Kuhn  \& Dr~A.~Madhavapeddy


% Main document

\section*{\bf Introduction, The Problem To Be Addressed}
PageRank (an algorithm which is used by Google to evaluate the ‘importance’ of a web page) is one of the most crucial fact

A problem of approximating algorithms which may be used by modern search engines is characterised by vast search space, wh

Firstly, there are little theoretical guarantees in this approach. Bounds, if any, referring to how much data needs to be

Another similar issue is referred to as ‘overfitting’: this describes a situation in which a classifier performs outstandi

Clearly, such limitations are hard to combat. Besides, machine learning techniques vary greatly, so clear and detailed fee

In this light, this project aspires to explore how machine learning techniques can be used to infer algorithms from search

\begin{figure}
\centering
\includegraphics[scale=0.5]{diagram1}
\caption{The training of the learner. The system enclosed within the dotted box will be implemented in this project. The m
\label{diag1}
```

```
\end{figure}

Most importantly, this approach gives me straightforward ways to reason about the performance of learning techniques

Evolving the search engine and, hence, the learner iteratively will result in comprehensive conclusions about the ef

\section*{\bf Starting Point}

\begin{itemize}
\item A project\cite{reid} was undertaken by the proposer, which developed a primitive algorithm to predict, given s
\item Python packages exist for manipulating web pages.
\item Wget is a Linux open source utility that can be used to clone web pages.
\item The paper describing PageRank is published and will be used to implement the algorithm.
\end{itemize}


\section*{\bf Resources Required}
\begin{itemize}
\item For this project I shall mainly use my own dual-core computer that runs Ubuntu Linux. I accept full responsibil
\item Backup will be to a Bitbucket repository and/or an external hard drive.
\item I will work on MCS computers should my main machine suddenly fail.
\end{itemize}
\section*{\bf Work to be done}

The project breaks down into the following sub-projects:

\begin{enumerate}
\item Decide on a category of search terms to explore in order to create a small network consisting of relevant web

\item Implement PageRank within this network.

\item Write a simple search engine incorporating PageRank and few other features.

\item Decide on the representation of the input for the learner and set up the framework to format it.

\item In advance set aside training and test data: this is necessary to then justify the evaluation of the classifie

\item Write a simple prototype for the learner\footnote {A Naive Bayesian would be a good prototype to use.} to test

\item Design, implement and test the learner.

\item Attempt to evolve the search engine to be more usable and complex and observe how the learner copes with the c


\end{enumerate}

\section*{\bf Success Criterion for the Main Result}


The project will be a success if...
\begin{itemize}
\item The resulting classifier can identify the importance of the PageRank factor in the given search engine.
\item The results of the experiment show how the chosen machine learning technique deals with various search engine
\end{itemize}
\section*{\bf Possible Extensions}
If I achieve my main result early I shall experiment with other machine learning techniques to see which perform bet
discovering dependencies between features of the page and its success in ranking results.
\section*{\bf Timetable: Work plan and Milestones to be achieved.}
```

```
Planned starting date is 19/10/2011.

\begin{enumerate}

\item {\bf 9 Oct - 19 Oct:}
\begin{itemize}
\item Do preliminary reading.
\item Familiarize myself with the field of machine learning.
\end{itemize}
{\bf Milestone: } Complete project proposal.
\item {\bf Oct 20 - Nov 3:}
    \begin{itemize}
    \item Decide which and how many websites should be cloned for use as the mini web.
    \item Prepare some training data and, separately, test data. This includes queries to be run on the search engine and
    \end{itemize}
\item {\bf Nov 4 - Nov 15:}
    \begin{itemize}
    \item Start writing a simple search engine and evaluate it on the test data.
    \end{itemize}
\item {\bf Nov 15 - Nov 25:}
    \begin{itemize}
    \item Finish the search engine.
    \item Start developing an early prototype for the learner.
    \end{itemize}
    {\bf Milestone: } Have a prototype of a complete system.
\item {\bf Nov 25 - Dec 15:}
\begin{itemize}
\item Evaluate the performance of the prototype learner.
\item Design and start implementing the final learner using the results obtained from the prototype as guidance.
\end{itemize}
\item {\bf Dec 16 - Jan 1:} Finish the implementation of the learner.
\item {\bf Jan 2 - Jan 16:}
     Evaluate the resulting classifier. Here is also good time to try a different design for the learner if the classifier
\item {\bf Jan 17 - Feb 1:} Start working on progress report.
{Milestone: } Write progress report.
\item {\bf Feb 2 - Feb 20} Implement extensions.

\item {\bf Feb 20 - Mar 5:} Evaluate extensions.

\item {\bf Mar 5 - Mar 25:} Write dissertation main chapters.

\item {\bf Mar 25 - April 10}  Further evaluation and complete dissertation.
{Milestone: } Dissertation final draft is finished.

\item {\bf April 11 - April 20:} Proof reading and submission.

\end{enumerate}

\begin{thebibliography}{9}

\bibitem{domingos}
  Pedro Domingos,
  \emph{A Few Useful Things to Know about Machine Learning},
  University of Washington.

\bibitem{reid}
  Alex Reid,
  \emph{Project 1},
  http://www.scienceforsearch.com/project1.asp.

\end{thebibliography}
```

Appendix B

Makefile

B.1 Makefile

```
# This is the Makefile for the demonstration dissertation
# written by Martin Richards
#
# Note that continuation lines require '\'
# and that TAB is used after ':' and before unix commands.

DISS = diss.tex refs.bib propbody.tex figs/diagram.eps makefile.txt

PROP = proposal.tex propbody.tex

help:
        @echo
        @echo "USAGE:"
        @echo
        @echo "make          display help information"
        @echo "make prop     make the proposal and view it using xdvi"
        @echo "make diss.ps  make a postscript version of the dissertation"
        @echo "make diss.pdf make a .pdf version of the dissertation"
        @echo "make gv       view the dissertation with ghostview"
        @echo "make gs       view the dissertation with ghostscript"
        @echo "make all      construct proposal.dvi and diss.ps"
        @echo "make count    display an estimated word count"
        @echo "make pub      put demodiss.tar on my homepage"
        @echo "make clean    remove all remakeable files"
        @echo "make pr       print the dissertation"
        @echo

prop:   proposal.dvi
        xdvi proposal.dvi

diss.ps:        $(DISS)
        latex diss
        bibtex diss
        latex diss
        bibtex diss
        latex diss
        bibtex diss
        dvips -Ppdf -G0 -t a4 -pp 0-200 -o diss.ps diss.dvi

diss.pdf:       diss.ps
        ps2pdf diss.ps

makefile.txt:   Makefile
        expand Makefile >makefile.txt
count:
        detex diss.tex | tr -cd '0-9A-Za-z \n' | wc -w

proposal.dvi: $(PROP)
        latex proposal

all:    proposal.dvi diss.ps

pub:    diss.pdf
        cp diss.pdf /homes/mr/public_html/demodiss.pdf
        make clean
        (cd ..; tar cfv /homes/mr/public_html/demodiss.tar demodiss)

clean:
        rm -f diss.ps *.dvi *.aux *.log *.err
        rm -f core *~ *.lof *.toc *.blg *.bbl
        rm -f makefile.txt

gv:     diss.ps
        ghostview diss.ps

gs:     diss.ps
        gs diss.ps

pr:     diss.ps
        lpr diss.ps
```

B.2 refs.bib

```
@BOOK{Lamport86,
TITLE = "{LaTeX} --- a document preparation system --- user's guide
and reference manual",
AUTHOR = "Lamport, L.",
PUBLISHER = "Addison-Wesley",
YEAR = "1986"}

@REPORT{Moore95,
TITLE = "How to prepare a dissertation in LaTeX",
AUTHOR = "Moore, S.W.",
YEAR = "1995"}
```

Appendix C

Project Proposal