|  |  |
| --- | --- |
| Glowing blue particles  Artificial Intelligence: Automatic News Extraction | Abstract  The documentation of an investigation into the automatic extraction of online news articles using an artificial intelligence and machine learning.  Ryder Franklin  BSc Computer Science: Final Year Project  Supervisor: Jun He |

#

# Abstract

## Acknowledgements

Jun He – for his continued support as my supervisor during this project.

Arnie (the cat) – for sitting patiently with me while I code.

## Contents

Table of Contents

[Artificial Intelligence: Automatic News Extraction 0](https://myntuac-my.sharepoint.com/personal/n0853033_my_ntu_ac_uk/Documents/N0853033_report.docx#_Toc89433910)

[Abstract 0](#_Toc89433911)

[Acknowledgements 1](#_Toc89433912)

[Contents 1](#_Toc89433913)

[Figures and Tables 1](#_Toc89433914)

[Chapters 2](#_Toc89433915)

[Chapter One: Introduction 2](#_Toc89433916)

[Chapter Two: Context 2](#_Toc89433917)

[Chapter Three: New Ideas 2](#_Toc89433918)

[Chapter Four: Implementation 3](#_Toc89433919)

[Chapter Five: Results and Discussion 3](#_Toc89433920)

[Chapter Six: Conclusions and Future Work 3](#_Toc89433921)

[References 3](#_Toc89433922)

[Bibliography 3](#_Toc89433923)

[Appendices 3](#_Toc89433924)

## Figures and Tables

# Chapters

## Chapter One: Introduction

This investigation is going to be researching, implementing and testing the usage of Artificial Intelligence to extract news articles from an internet source, with the topic selected by the user. Similar to a search engine, the functionality of the program will be based in collecting articles based upon a given topic. However, my project will be diverging from a search engine in that it will only collect news sources and, specifically, news from reliable sources. These sources may be counted as reliable due to their quality of fact checking, their reputation for upholding impartiality or their frequency of updating news.

## Chapter Two: Context

Some articles to finish reading and analyse –

* <https://dl.acm.org/doi/pdf/10.1145/3184558.3191600>
  + Article about a search engine that uses AI to search large amount of data
* <https://ojs.aaai.org/index.php/aimagazine/article/view/2601>
  + A book about an AI that is used in a digital library search engine
    - Check ntu library for a physical copy
* <https://www.igi-global.com/chapter/content/64436>
  + A book about AI in search engines, noted as Artificial Intelligence Enabled Search Engines (AIESE)
    - Check ntu library for a physical copy

## Chapter Three: New Ideas

\*rewrite this more formally

My newest idea for this project is the creation of a machine learning algorithm that ranks an articles relevance to the provided user input using various factors and supplies the user with the “most” relevant articles, according to the algorithm.

Another idea is the creation of a machine learning algorithm that ranks a websites’ chance of producing “fake” or altered news articles, which may be biased or irrelevant to the users’ input. For example, a website such as ScienceDaily (cite) that posts purely scientific research that is backed up and authenticated by multiple people, are more likely to produce unbiased and provable news articles. This will be especially important when collecting news articles on current events, especially those with political undertones, that an author may wish to write about in a way that drives a political agenda.

A historical example of news sources potentially being unreliable may be the Hillsborough disaster in 1989 and the articles that followed by various newspapers, most notably The Sun. The disaster was a fatal “human crush” that occurred during a football match at the Hillsborough Stadium in Sheffield in the April of 1989.

Based on false information provided by a police inspector, unnamed police officers and a local Conservative MP, The Sun wrote various articles claiming that both the victims and those who witnessed this event had acted as “hooligans” and disrespected those who required medical attention, even claiming that they “picked the pockets” of victims (cite this). Although this information was later proved to be false, The Sun newspaper gained a reputation of being an unreliable news source within Liverpool, who’s football team were playing at Hillsborough stadium at the time, and a ban of all journalists from The Sun from entering Liverpool Football Clubs’ grounds was issued in 2017. (cite this)

* Could this be more suited in context? Although it is justifying why news sources should be checked for reliability. Hmm.

## Chapter Four: Implementation

## Chapter Five: Results and Discussion

## Chapter Six: Conclusions and Future Work

# References

\*format this in refworks

Chapter One:

Chapter Two:

<https://dl.acm.org/doi/pdf/10.1145/3184558.3191600>

<https://ojs.aaai.org/index.php/aimagazine/article/view/2601>

<https://www.igi-global.com/chapter/content/64436>

Chapter Three:

sciencedaily.com

https://www.bbc.co.uk/news/uk-england-merseyside-29564707

https://www.bbc.co.uk/news/uk-england-merseyside-38933817

# Bibliography

# Appendices