Abstract

**A compulsory section, of at most 300 words**

This section should precis the project context, aims and objectives, and main contributions (e.g., deliverables) and achievements; the same section may be called an abstract elsewhere. The goal is to ensure the reader is clear about what the topic is, what you have done within this topic, and what your view of the outcome is.

The former aspects should be guided by your specification: essentially this section is a (very) short version of what is typically the first chapter. Note that for research-type projects, this must include a clear research hypothesis. This will obviously differ significantly for each project, but an example might be as follows:

My research hypothesis is that a suitable genetic algorithm will yield more accurate results

(when applied to the standard ACME data set) than the algorithm proposed by Jones and

Smith, while also executing in less time.

The latter aspects should (ideally) be presented as a concise, factual bullet point list. Again, the points will differ for each project, but it might be as follows:

* I spent 120 hours collecting material on and learning about the Java garbage-collection sub-system.
* I wrote a total of 5000 lines of source code, comprising a Linux device driver for a robot (in C) and a GUI (in Java) that is used to control it.
* I designed a new algorithm for computing the non-linear mapping from A-space to B-space using a genetic algorithm, see page 17.
* I implemented a version of the algorithm proposed by Jones and Smith in [6], see page 12, corrected a mistake in it, and compared the results with several alternatives.

# Dedication and Acknowledgements

**A compulsory section**

It is common practice (although totally optional) to acknowledge any third-party advice, contribution or

influence you have found useful during your work. Examples include support from friends or family, the

input of your Supervisor and/or Advisor, external organisations or persons who have supplied resources

of some kind (e.g., funding, advice or time), and so on.Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Taught Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, this work is my own work. Work done in collaboration with, or with the assistance of others, is indicated as such. I have identified all material in this dissertation which is not my own work through appropriate referencing and acknowledgement. Where I have quoted or otherwise incorporated material which is the work of others, I have included the source in the references. Any views expressed in the dissertation, other than referenced material, are those of the author.

Daniel Page, Wednesday, 16 March 2022

Table of Contents

[Dedication and Acknowledgements ii](#_Toc98342026)

[List of Figures v](#_Toc98342027)

[List of Tables vi](#_Toc98342028)

[Summary of Changes viii](#_Toc98342029)

[Supporting Technologies ix](#_Toc98342030)

[Notions and Acronyms x](#_Toc98342031)

[Chapter 1: Introduction 1](#_Toc98342032)

[Chapter 2: Background 5](#_Toc98342033)

[Chapter 3: Project Execution 13](#_Toc98342034)

[3.1. Example Section 13](#_Toc98342035)

[3.1.1. Example sub-section 14](#_Toc98342036)

[Chapter 4: Critical Evaluation 17](#_Toc98342037)

[Chapter 5: Conclusion 18](#_Toc98342038)

[Bibliography 19](#_Toc98342039)

[Appendix A: An Example Appendix 20](#_Toc98342040)

# List of Figures

[Figure 1: This is an example figure caption 14](#_Toc98342041)

# List of Tables

[Table 1: Example table caption 16](#_Toc98342047)

Ethics Statement

**A compulsory section**

In almost every project, this will be one of the following statements:

* “This project did not require ethical review, as determined by my supervisor, [fill in name]”; or
* “This project fits within the scope of ethics application 0026, as reviewed by my supervisor, [fill in name]”; or
* “An ethics application for this project was reviewed and approved by the faculty research ethics committee as application [fill in number]”.

See Section 3.2 of the unit Handbook for more information. If something went wrong and none of those three statements apply, then you should instead explain what happened.

# Summary of Changes

**A conditional section**

If and only if the dissertation represents a resubmission (e.g., as the result of a resit), this section is compulsory: the content should summarise all non-trivial changes made to the initial submission. Otherwise, you can omit it, since a summary of this type is clearly nonsensical.

When included, the section will ideally be used to highlight additional work completed, and address criticism raised in any associated feedback. Clearly it is difficult to give generic advice about how to do so, but an example might be as follows:

* Feedback from the initial submission criticised the design and implementation of my genetic algorithm, stating “there seems to have been no attention to computational complexity during the design, and obvious methods of optimisation are missing within the resulting implementation”. Chapter 3 now includes a comprehensive analysis of the algorithm, in terms of both time and space. While I have not altered the algorithm itself, I have included a cache mechanism (also detailed in Chapter 3) that provides a significant improvement in average run-time.
* I added a feature in my implementation to allow automatic rather than manual selection of various parameters; the experimental results in Chapter 4 have been updated to reflect this.
* Questions after the presentation highlighted a range of related work that I had not considered: I have made a number of updates to Chapter 2, resolving this issue.

# Supporting Technologies

**An optional section, of at most 1 page**

This section should present a detailed summary, in bullet point form, of any third-party resources (e.g., hardware and software components) used during the project. Use of such resources is always perfectly acceptable: the goal of this section is simply to be clear about how and where they are used, so that a clear assessment of your work can result. The content can focus on the project topic itself (rather, for example, than including “I used Word to prepare my dissertation”); an example is as follows:

* I used the Java BigInteger class to support my implementation of RSA.
* I used a parts of the OpenCV computer vision library to capture images from a camera, and for various standard operations (e.g., threshold, edge detection).
* I used an FPGA device supplied by the Department, and altered it to support an open-source UART core obtained from <http://opencores.org/>
* The web-interface component of my system was implemented by extending the open-source WordPress software available from <http://wordpress.org/>

# Notions and Acronyms

**An optional section**

Any well written document will introduce notation and acronyms before their use, *even if* they are standard in some way: this ensures any reader can understand the resulting self-contained content.

Said introduction can exist within the dissertation itself, wherever that is appropriate. For an acronym, this is typically achieved at the first point of use via “Advanced Encryption Standard (AES)” or similar, noting the capitalisation of relevant letters. However, it can be useful to include an additional, dedicated list at the start of the dissertation; the advantage of doing so is that you cannot mistakenly use an acronym before defining it. A limited example is as follows:

|  |  |  |
| --- | --- | --- |
| AES | : | Advanced Encryption Standard |
| DES | : | Data Encryption Standard |
|  | : | The Hamming weight of |
|  | : | A finite field with elements |
|  | : | The -th bit of some binary sequence st. |

# Chapter 1: Introduction

**Unlike the frontmatter up to and including the Summary of Changes, which you should not deviate from, Chapters 1–5 represent a suggested outline only. This outline will only be appropriate for a specific type of project. You should talk with your supervisor about the best way to structure your own dissertation, but ultimately the choice is yours. However, almost every project will want to include the content discussed in these chapters in some way. For more advice on structuring your dissertation, see the unit handbook.**

This chapter should introduce the project context and motivate each of the proposed aims and objectives. Ideally, it is written at a fairly high-level, and easily understood by a reader who is technically competent but not an expert in the topic itself.

In short, the goal is to answer three questions for the reader. First, what is the project topic, or problem being investigated? Second, why is the topic important, or rather why should the reader care about it? For example, why there is a need for this project (e.g., lack of similar software or deficiency in existing software), who will benefit from the project and in what way (e.g., end-users, or software developers) what work does the project build on and why is the selected approach either important and/or interesting (e.g., fills a gap in literature, applies results from another field to a new problem). Finally, what are the central challenges involved and why are they significant?

The chapter should conclude with a concise bullet point list that summarises the aims and objectives. For example:

The high-level objective of this project is to reduce the performance gap between hardware and software implementations of modular arithmetic. More specifically, the concrete aims are:

1. Research and survey literature on public-key cryptography and identify the state of the art in exponentiation algorithms.
2. Improve the state-of-the-art algorithm so that it can be used in an effective and flexible way on constrained devices.
3. Implement a framework for describing exponentiation algorithms and populate it with suitable examples from the literature on an ARM7 platform.
4. Use the framework to perform a study of algorithm performance in terms of time and space, and show the proposed improvements are worthwhile.
   1. Motivation

The judgment is a key to understand the legal system and is the final output of the judicial system's handling of a case. It is not only for the parties to see, but also open to the whole society which shows how a specific case was handled and suggests how similar cases will be handled. However, legal judgments in the UK are usually lengthy and complex documents, full of technical and intricate legal arguments, all of which take a long time to read once there is a need to find similar arguments to support.

Unlike other countries where cases are heard by a single judge, the UK is different in that cases heard by UK House of Lords (UKHL) involve a panel of judges considering the issues. The number of judges on the panel is usually an odd number to ensure that there is a clear majority result. After deliberation, the judges discuss and debate until a consensus is reached, and then one or the panel works together to draft a judgment. This process ensures that the final decision is reasonable, fair and represents a unified approach by UKHL. After 2009, Supreme Court (UKSC) is created to replace UKHL and made changes to the structure of the judgement document. There are situations that judges quoting each other and expressing opinions based on other judges' viewpoints rather than independently in dozens of pages of reports.

Despite their importance, it is still difficult to obtain concise summaries of these decisions, making effective legal learning and comprehension less easy. By visiting the British and Irish Legal Information Institute (BAILII) website, it does not contain summaries of UKHL cases prior to 2009, so when we need to understand the content of a specific case, it is time-consuming to read the full document. While the official website of UKSC provides a summary which only contains part of the content like issues and facts, there are only decided cases which also provide a press summary with a uniform layout on National Archives, and its page count is mostly 4-6 pages but maybe not suitable for citizens to understand.

Currently, all UK case decisions are manually summarized, which means that all of this content on the website needs the institution to find professionals to processed, which is not only labour-intensive, but may be biased by different people dealing with different standards, and there are no tools available that can provide summaries immediately after the decision or provide customized summaries to meet individual needs. This paper attempts to solve this problem by using an auto summarization system specifically designed for British legal decisions to save time for those who manually prepare case summaries or have needs to access it.

Valvoda et al. introduce Agreement Statements and Majority Opinion (ASMO) which provides us with a way to automatically identify two elements from UKHL cases. They set up different rules and train the model to determine whether the sentence contains AS and whether the two sides of the opinion completely agree with each other or have independent views. It finally return the judge who has the majority opinion in the case as the main focus of our final summary.

In addition, Hachey and Grover's SUM system uses machine learning to identify rhetorical tags in the sentence which categorising legal texts into 7 categories and whether the sentence is relevant to the judgement. It outputs the relevance of each sentence as a rank score which is an important parameter for the final summary output. The system was further improved by Imansyah and Amy's further research such as using Conditional Random Fields (CRFs) to improve the rhetorical and relevance classifiers, adding Recall-Oriented Understudy for Gisting Evaluation (ROUGE) metrics to the output of the system etc. Amy's SUMO system integrates the above researches, use ASMO results as a feature greatly improves the quality of extracted summaries by extensively evaluating system-generated summaries of HOJI datasets before 2003.

We will integrate the SUMO and ASMO pipelines so that the new system can be applied to data other than annotations pre labelled by professionals, including UKHL and UKSC data. We will make sure it can dynamically process cases based on user input and output summaries of varying complexity as required. This system, which highlights important facts from judgments as well as precedents, will save time and reduce human error for those who manually write case summaries. We also hope to enable real-world deployment of our system, improving accessibility for legal professionals, law students, and UK citizens with varying needs for the law.

在英国，法律判决是理解法律体系的关键，并且是司法系统对案件处理的最终产出。它不仅供当事人查看，还向整个社会开放，展示了特定案件的处理方式，并暗示了类似案件将如何处理。然而，英国的法律判决通常是冗长而复杂的文件，充满了技术性和错综复杂的法律论点，一旦有需求去寻找相似论点，都需要花费长时间进行阅读。

在其他国家，案件的审理基本由一位法官进行，而英国不同的点在于，UKHL审理的案件涉及一组法官对问题进行审议。小组法官的数量通常是奇数，以确保有一个明确的多数结果。经过审议后，法官们进行讨论和辩论，直到达成一致意见，然后由其中一位或小组共同起草判决。这个过程确保了最终的决定是合理的、公正的，并代表了UKHL的统一态度。在2009年之后，UKSC的出现取代了UKHL并对judgement文档的结构做出了修改。几十页的judgement会出现法官之间互相引用并且表达基于以上法官的意见而不是互相独立。

尽管法律判决具有重要意义，但要获取这些裁决的简明摘要仍然很困难，这使得有效的法律学习和理解变得不那么容易。通过访问Bailli网站，我们无法查找到关于2009年之前UKHL的case的摘要，所以当我们需要了解内容时，需要阅读完整的文档十分的耗费时间。而在UKSC的官网提供了只包含部分内容的summary，有一些case也提供了统一排版的press summary，其页数也多为4-6页。目前，所有英国的案件判决都是手动摘要的，意味着网站上的这些内容均需要机构安排专业人员进行处理，不仅耗费人力，可能会有不同人员处理标准不同的问题产生偏差，并且没有可用的工具可以在裁决后立即提供摘要或提供符合个人需求的定制。这种缺乏使得阅读冗长的法律文件并及时获取重要的法律信息变得更加困难。这篇paper主要是想解决这个问题，使用专门针对英国法律判决的摘要系统，为手动编写案例摘要的人节省时间。

Valvoda等人之前的研究ASMO为我们提供了一种自动获取多数意见的方法，他们设置不同的rule并且训练模型判断句子中是否包含AS，并且判断观点双方是否完全认同或者拥有独立观点，最后返回在case里本人卫视多数意见的法官作为我们最后summary的主要关注对象。

除此之外，Hachey和Grover的SUM系统使用机器学习来识别判决句子中的修辞标签以及句子是否与正文相关，并且为每个句子输出相关性作为rank 分数，它的排序是最后输出summary的重要参考标准。该系统经过Imansyah 和 Amy的进一步研究比如使用使用条件随机场改进修辞和相关性分类器，添加Recall-Oriented Understudy for Gisting Evaluation（ROUGE）指标系统输出等有进一步的提升。Amy 的SUMO system整合了以上两个研究成功，将ASMO结果作为参考依据，通过对系统生成的2003年以前的HOJI数据集摘要进行广泛评估，极大提高了提取式摘要的质量。

我们将整合SUMO和ASMO pipeline，让新的系统可以适用于专业人士Annotation以外的数据，包括2009年以前UKHL和迁移到UKSC之后的数据，可以根据用户的输入动态处理案件，并根据需求输出不同复杂程度的摘要。这个系统可以高亮判决中重要的事实以及判例，将为手动编写案例摘要的人节省时间，减少人为错误。我们还希望能够实现我们的系统的现实世界部署，提高法律专业人士、法学生和英国公民等对法律有不同需求的人的便利性。

* 1. Aim and Hypotheses

In this project, we are working on the development of an automated summarization system combining ASMO and SUMO, which is able to automatically extract key features such as contextual logic between different judges based on user inputs. It can also automatically analyze the structure of original UKSC and UKHL judgement from different years on the Bailli website, including the paragraph numbers and the names of the speakers etc., and generate summaries that meet the requirements of both official formats and user needs. The official templates are those that we have gathered the style of the press summaries that are accessible on the UKSC website and the handwritten summaries that are published by the Law Reporting Council of England and Wales (ICLR).

To achieve these aims, we make the following hypotheses:

* Rewrite some of the ASMO and SUMO code to utilize the latest Python libraries as there is currently an inconvenient conflict due to the fact these different modules use two incompatible versions of library. This will successfully ensure that all the functionality can run smoothly with original idea.
* Replace some of the features generators by a no-longer maintained since spacy 2.1.9 parser called Blackstone used in SUMO with features generated by a maintained parser such as spaCy. And we retrain the model used in Amy and Valvoda's system with new feature set will have similar accuracy and results or slightly effect.
* The new system integrates the two pipelines and accesses the Bailli website, obtains the same content before 2003 as the ASMO and HOLI corpus but without professional annotation processing, stores ASMO and SUMO feature values in the same file, and verifies that an unprocessed textual data achieves the same results.
* The new system uses BeautifulSoup to analyze the structure of pre-2009 UKHL and post-2009 UKSC text to be possible summarise more recent cases and test how well system performs on cases outside of the data-set by ROUGE metrics and user evaluation, and test whether they are similar to the manual summaries available on the ICLR and UKSC websites.

在这个项目中，我们致力于开发一个结合了ASMO和SUMO的自动摘要系统，该系统能够根据用户的输入自动提取其中重要的特征值比如上下文的引用逻辑等，能够自动分析bailli网站不同年份的UKSC和UKHL的judgement report的行文结构包括段落号码和发言法官的名字等，能够生成符合用户要求和官方模板的摘要。官方模板指的是我们收集的英格兰和威尔士法法律报告委员会（ICLR）发布的手动撰写的摘要以及UKSC网站上提供的press summary的格式。

为了成功实现这一目标，我们提出以下假设：

* 统一环境中的需要的package的版本，修改部分代码可以成功解决ASMO和SUMO中关于使用的库版本不同的问题，保证所有的功能可以顺利运行。（rewriting some of the ASMO and SUMO code to utilise the latest Python libraries (as there is currently an inconvenient conflict due to the fact these different modules use two  incompatible version of library - even though the changes are relatively cosmetic)
* 替换掉SUMO中使用的在spacy 2.1.9之后没有再进行维护的blackstone，并且用现在还在不断更新的spacy或其他模型替代，重新训练Amy 和 Valvoda的代码中所使用的模型，并达到类似的准确率和 效果（replacing some of the features generators by a no-longer maintained parser called Blackstone with features generated by a maintained parser such as spaCy）
* 新系统将两个pipeline结合起来起来并且接入Bailli网站，获取其中的2003年前和ASMO语料库一样但是未经annotation处理的内容，将ASMO和SUMO的特征值储存到同一个文件中，并且验证一个未经处理的文本数据可以达到一样的效果
* 新系统使用BeautifulSoup分析2009年以前UKHL和2009年之后UKSC的文本结构，生成更加准确的摘要，通过ROUGE指标和用户评估进行衡量摘要的质量，检测它是否与ICLR和UKSC网站上的手动摘要类似
  1. Contribution
  2. Layout

# Chapter 2: Background

This chapter is intended to describe the background on which execution of the project depends. This may be a technical or a contextual background, or both. The goal is to provide a detailed explanation of the specific problem at hand, and existing work that is relevant (e.g., an existing algorithm that you use, alternative solutions proposed, supporting technologies).

Per the same advice in the handbook, note there is a subtly difference from this and a full-blown literature review (or survey). The latter might try to capture and organise (e.g., categorise somehow) *all* related work, potentially offering meta-analysis, whereas here the goal is simple to ensure the dissertation is self-contained. Put another way, after reading this chapter a non-expert reader should have obtained enough background to understand what *you* have done (by reading subsequent sections), then accurately assess your work against existing relevant related work. You might view an additional goal as giving the reader confidence that you are able to absorb, understand and clearly communicate highly technical material and to situate your work within existing literature.

Since our system applies machine learning and natural language processing to the legal domain, so understanding the legal context is critical to our next automation tasks. In this chapter, we explores the legal and technical background necessary to bring together our efforts to create an automated summarization system for UKHL and UKSC judgments.

我们研究的是将ML和NLP应用在法律领域的新系统，所以了解法律环境对于我们接下来的自动化任务至关重要。 在本章中，我们深入研究了必要的法律和技术背景，以将我们为 UKHL 和 UKSC 判决创建自动汇总系统的努力结合起来。

1. **Legal Background**

In the legal field, the principles of the common law system are the foundation for making legal decisions, and the retrieval and induction of precedents is the basis for understanding common law. In addition, the evolution of the English judicial system, particularly the transition from the House of Lords to the Supreme Court, has had a significant impact on legal practice and research. We explore these fundamental aspects and provide insights into accessing judgment through platforms such as Bailii and official websites.

在法律领域，普通法体系的原则是做出法律决定的基石，判例的检索和归纳是理解英美法的基础。 此外，英国司法体系的演变，特别是从上议院到最高法院的过渡，对法律实践和研究具有重大影响。 我们探索这些基本方面，并提供通过Bailli 和官方网站等平台获取判断的见解。

**i. The Common Law**

* Common law, also known as case law is one of the two major legal systems along with the civil law in the world. It differs most from the use of statutory law in that it follows precedent where the judge applies the written law to a specific case and makes a judgment within statutory law. The common law does not have a written law, so the judge can only rely on previous precedents to rule on the current case. It is its most notable feature that it always cited a large number of precedents to explain the legal provision. The process of argumentation in the judgment of a higher court is binding authority for the lower court, and the lower court cannot go against it. Even courts that do not have a superior-subordinate relationship, the content of their judgments have a general persuasive effect and non-binding authority on each other, and do not necessarily have to comply with, but can be used to support their own arguments.
* Common law judges gradually developed common law rules as they deal with cases, but these rules were not explicitly written out like enactments but were hidden in the case. Commentaries on the Laws of England, written by Sir William Blackstone, which became the best-known description of the English common law. He organized more than 2000 cases into categories which give insights to judges what the law is and ho to deal with new case. But in the process of implementing common law, two problems are usually faced: 1. which one or which of the many precedents should the judge follow to make a decision, and the lawyer could use as his or her own argument to support the their own position; 2. how to adapt to the significant changes in society and periods if they are followed the precedents strictly.
* A judgment contains a lot of content, such as the facts of the case, the reasons for the judgment and references to previous cases, etc. However, the reasons will not be explicitly stated for the judgment like writing in a specific section, the reader needs to read through it to summarised and find out. Different judges would have a different interpretations of the case, leading to different perspective in new cases. So reading judgements is something that all people involved common law need to do. Law students need to study a lot of cases to understand legal principles and precedents for their future careers. Lawyers and legal professionals also need to draft legal documents and provide legal advice to their customers with sufficient amount of cases in mind.
* 英美法系又名普通法，他与制定法在使用上最大的差别是遵循先例。使用制定法时，法官将成文的法条适用到具体的案件中，做出判决。而普通法没有成文的法条，因此法官只能依据之前的判例来裁决当下案件，这就是他最突出的特点即引用大量的判例来解释法律条文.上级法院的判决书里的论证过程对下级法院来说具有约束力Binding Authority，下级法院不能与之相违背。即便是没有上下级关系的法院，其判决书内容对彼此也有一般的说服力，不一定要遵守，但是可以用来支持自己的论证。
* 普通法法官在处理纠纷的过程中逐步发展出普通法规则，但是这些规则不像制定法那样明确写出来而是隐藏在案件里。大法官布雷克顿写的《英格兰的法律与习惯》奠定了普通法的基础，对国王两千多个案例分门别类的梳理，有助于法官从中发现法律。但是在common law实施的过程中通常面临两个问题，1.那么多先例，法官应该遵循哪一个或者哪些做出判决，律师应该使用哪些先例作为自己的论据。2.如果严格遵循，如何应对时代和社会的巨大变化。
* 一个judgement包含很多内容，比如案件事实，判据理由和对过往案例的引用等，但实际上一个案件不会明确的说出判据理由,而是隐藏在法官的整个推理过程中，需要读者去阅读梳理才能发现，而不同的法官可能会有不同的理解，导致对不同新案件的看法也不同。所以阅读大量judgement是所有与common law相关的人需要做到的。所以这个法律背景下，法学院的学生需要研究大量案例来理解法律原则和先例，为他们未来的职业生涯打下基础。律师和法律专业人士也需要通过这个举动为委托人辩护、起草法律文件和提供法律咨询。

**ii. UKHL and UKSC Judgments**

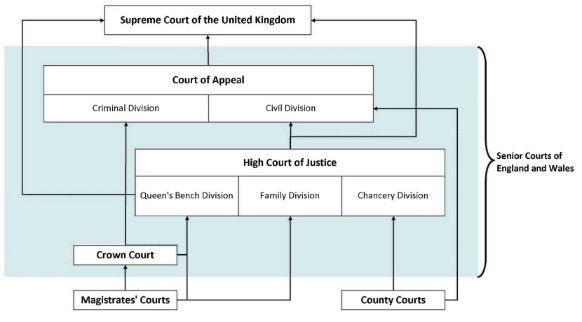
A key component of the Constitution is the courts, which serve as the last arbitrator between citizens and the State. Before the Supreme Court was founded in 2009, the highest appeal court in United Kingdom is the House of Lort which was established on 14th century. It used to have the authority to make the final decision in all civil cases in UK, as well as all criminal cases in the country except the one in Scotland. However, its jurisdiction was not shared by the whole house but by a committee of "Law Lords", legally experienced member in UKHL who had exclusive authority to hear appeals from subordinate courts. Under the Appellate Jurisdiction Act 1876 the twelve Lords of Appeal in Ordinary are appointed to manage the complex judicial business of the House. They are responsible for hearing a wide range of complex legal cases, including those involving significant legal principles and the public rights. Judicial functions may also be exercised by the Lords of Appeal, the other members of the House who have held high judicial office.

Not every cases will be heard by all twelve law lords but by Appellate Committees after World War II. The Senior Lord selected the members of the committees which are usually made up of five judges from UKHL, but when reviewing significant cases, the committee may have more than five members. However, as it serves as a legislative body as well as a court, it creates potential conflicts of interests and affects the independence of the judicial independence. Therefore, the judicial duties of UKHL were officially abolishes in 2009 and the relevant functions were transferred to Supreme Court of UK which continues to perform its role as the final court of appeal and make decisions in cases involving great public significant.

In most cases, UKSC will hear arguments from multiple justices and the number of judges is usually being odd to guarantee a majority opinion at the end. They will only write decisions and judgement documents at the end when they come up with a consensus. Generally, cases will be heard by five judges and this group will be enlarged when the case is particular serious. To data, only two cases have been considered by 11 justices so far, both involving matters of major constitutional importance: R (Miller) v Secretary of State for Exiting the European Union (2016 debate, 2017 decision) and R ( Miller v The Prime Minister and Cherry v Advocate General for Scotland (argued and decided 2019).

法院是公民与国家之间的最终仲裁者，因此是宪法的基本支柱。在2009年最高法院成立之前，英国的最高上诉法院始创于14世纪的上议院，它曾经拥有对联合王国内所有民事案件及除苏格兰以外的刑事案件最终审判的权力。但是它的法职能并不由全院共同行使，而是交由院内具法律经验的议员们，即上议院高等法官，仅能审理来自下级法院的上诉案件。依《司法上诉受理法令》（Appellate Jurisdiction Act）的规定，一共指派的12名常任上诉法官管理院中繁杂的司法事务，负责审理各种复杂的法律案件，包括涉及重大法律原则、公共利益或有重大影响的案件。一般上诉法官（Lords of Appeal，院中其他具高阶司法职位的议员）也可行使司法职能。

并不是每宗案件都会由所有上议院高等法官会审，第二次世界大战以后，上诉案由各受理上诉委员会（Appellate Committees）专责审理，委员会一般由五名上院法官组成，成员由首席上院法官选定。审议重大案件时，受理上诉委员会成员人数可多于五名。但其作为立法机构和法院的双重角色造成了潜在的利益冲突，并破坏了司法机构的独立性。所以在2009年上议院的司法职能被正式废除，有关职能转移到新成立的[联合王国最高法院](https://zh.wikipedia.org/wiki/%E8%81%AF%E5%90%88%E7%8E%8B%E5%9C%8B%E6%9C%80%E9%AB%98%E6%B3%95%E9%99%A2)，继续履行其作为终审上诉法院的作用，对具有重大公共意义的案件作出裁决。一般而言，最高法院审理的案件会由多名法官会审，会审人数多为单数，以确保有大比数的裁决。会审后法官达成一致共识，才会撰写判辞。一般案件会由5名法官会审，而特别重大的案件会由9位法官会审。迄今为止，只有11位大法官审理过两次案件（均涉及重大宪法重要性事项）：[R （Miller） v Secretary of State for Exiting the European Union](https://en.wikipedia.org/wiki/R_(Miller)_v_Secretary_of_State_for_Exiting_the_European_Union" \o "R (Miller) v Secretary of State for Exiting the European Union) 一案（2016年辩论，2017年裁决）和[R （Miller） v The Prime Minister and Cherry v Advocate General for Scotland](https://en.wikipedia.org/wiki/R_(Miller)_v_The_Prime_Minister_and_Cherry_v_Advocate_General_for_Scotland" \o "R (Miller) v The Prime Minister and Cherry v Advocate General for Scotland)一案（2019年辩论和裁决）。



1. **Bailli**:

As the core institution of the British legal system, the decision of UKHL and UKSC are crucial to guiding legal practice and defending civil rights. When we implement the auto summarize task of judgements, we need reliable data sources to gather these documents, also a easier and more suitable way for user to use in the future. We selected BAILII, a free and extensive database that offers essential legal resources in the UK and Ireland, to increase the popularity and accessibility of the judgements.

BAILII was established by various organizations including Society for Computers & Law, Chairman Neil Cameron, Lord Justice Brooke. They were concerned about the difficulty of common court users to access court decisions and were inspired by were inspired by the Australasian Legal Information Institute (AustLII) LII. In August 2019, BAILII included 102 databases covering 10 jurisdictions. Through the Bailli website, we can easily obtain the judgment documents of UKHL and UKSC and use them for the development of our subsequent research codes and data analysis.

作为英国法律体系的核心机构，最高法院的裁决对于引导法律实践和维护公民权利至关重要。在研究英国高等法院（UKHL）和最高法院（UKSC）的判决时，我们需要可靠的数据来源来获取这些判决文书。而为了使这些判决更加普及和易于获取，我们选择了BAILII这个提供英国和爱尔兰主要法律材料的免费且全面的数据库，其收录了大量的法律文书，包括UKHL和UKSC的判决。

BAILII是由包括计算机与法律协会（Society for Computers & Law）在内的各种活动家经过长期而艰苦的运动后成立的，例如当时的主席尼尔·卡梅伦（Neil Cameron），大律师劳里·韦斯特·奈特斯（Laurie West-Knights QC），萨维尔勋爵（Lord Saville）和布鲁克大法官（Lord Justice Brooke），他们担心普通法院用户无法获得法院判决，并受到澳大利亚法律信息研究所（AustLII）LII的启发。[3]其目的是提供对公开法律信息的免费访问。[1] 2006年，BAILII纳入了来自5个司法管辖区的14个数据库。[4] BAILII网站由伦敦大学高级法律研究所和科克大学学院法学院（UCC）联合主办。

通过Bailli网站，我们可以方便地获取UKHL和UKSC的判决文书，并将其用于我们后续研究代码的开发和数据分析。公众可以方便地获取到UKHL和UKSC判决的对应内容，从而更方便地使用我们系统了解案件的审理过程和法官的裁决理由。

**iii. Structure of Judgments**

## As there are so many UKHL and UKSC cases, we have to understand what each section of the law report means when we make the system automatically create summaries. The structure and components of the report have become one of our crucial foundations. The entire automated process does not involve the pre-procession or annotation by other human staffs, we need to use algorithms to extract and analyse key elements from the lengthy documents and group them into the feature set of our pipeline.

The Incorporated Council of Law Reporting (ICLR) is a registered charity which publishes legal reports on English law. According to the practice direction issued by Lord Judge when he was Lord Chief Justice of England and Wales, the Law Reports are "the most authoritative reports" and should always be "cited in preference where there is a choice.", which is published by ICLR. Here is the report of case [2002] UKHL 9 (28th February, 2002) on the BAILII website.

All legal report contains certain important parts such as title, court of appeal, etc. We need to pay attention to the following information when we analyse:

1. Title section usually include the names of both defendants, as shown in figure, it’s ‘National Westminster Bank Plc v. Amin and Another’. The Latin word "queen" (Regina) or "king" (Rex) is typically used as the criminal cases are usually prosecuted by Crown(Regina v. Mirza (Appellant) (On Appeal from the Court of Appeal (Criminal Division)). When a case only involve one party, ‘in re’(in matter of) is placed before the case name such as bankruptcy cases (In re Gallagher)or use ‘ex rel when the case involves extraordinary judicial relief.
2. The judges who took part in the decision are important as the final result of the cases is decided by five of them. We need to identify the corresponding content for each speaker in entire documents. As shown in figure, the name of judge is written before the first paragraph of each new speech. Each paragraph has a number at beginning which indicates the position of the content and also make it easier to refer to when they have considerations related to.
3. At the end of each speeches, it will mark the specific decision the committee make, allowing or rejecting the plaintiff’s claim. In this case, every judge allow the appeal and the Lord Scott of Foscote gave further explanation and arguments to describe the processing.
4. The most important part of the whole documents is the judgement of the case which including the background, fact and reason for appeal, and also the case stated by a lower court on an appeal by case stated. In the committee, there will be one or more than one judges provide a majority opinion which cited precedents and reasons on behalf of others. in addition to our example that each judge has their own section to speak, there’s situation that several judges collaborated to speak, that is, they work together to construct a judgement that in one section. This shows a type of fully agreement between the partners, like in case, they mentioned ‘LADY HALE, DELIVERING THE JUDGMENT OF THE COURT’. These different format also give us another challenge that which is the majority opinion as it’s definition is the line of reasoning agreed by more than half of the judges. we will explain the algorithm of it in next section.

## 面对海量的UKHL和UKSC判决书，当我们自动生成摘要时，我们需要明白law report中的每一部分代表了什么内容，他的结构和组成部分成了我们的重要指标之一。整个自动化的过程不涉及额外工作人员的处理和标注，所以利用算法来提取和分析这些判决文书中的关键要素，并将其纳入我们的整个pipeline的特征集中，是自动摘要系统的开发重要的基础。

## 英格兰和威尔士法律报告委员会 （ICLR） 是一家位于[英国](https://en.wikipedia.org/wiki/England" \o "England)[伦敦](https://en.wikipedia.org/wiki/London)的[注册慈善机构](https://en.wikipedia.org/wiki/Charitable_organization)，负责发布[英国法律](https://en.wikipedia.org/wiki/English_law)[的法律报告](https://en.wikipedia.org/wiki/Law_report)。ICLR发布的主要系列报告是《[法律报告](https://en.wikipedia.org/wiki/Law_Reports" \o "Law Reports)》，根据[法官](https://en.wikipedia.org/wiki/Igor_Judge,_Baron_Judge)在担任英格兰及威尔士首席大法官期间发出的实务指示，《法律报告》是“'最权威的报告'，应始终'在有选择的情况下优先引用'。图一中显示的是在BAILII网站上[2002] UKHL 9 (28th February, 2002)这个案件的judgement。

任何法律报告，都包含某些重要的部分比如标题，上诉法院等，我们需要关心的以下内容

1. title部分一般被告双方的名字组成, 在这个例子中，National Westminster Bank Plc v. Amin and Another. 而刑事案件通常以皇室提起，using the Latin for “queen” (Regina) or “king” (Rex)来代替(**Regina v. Mirza (Appellant) (On Appeal from the Court of Appeal (Criminal Division))**)。当只涉及一方当事人的案件，一般将in re（in the matter of） 加在案件名称前，如破产案件、监护人案件等。（Judgments - In re Maye (AP) (Appellant) (Northern Ireland) [2008] UKHL 9）或者使用 ex rel当案件涉及非正常司法救济的特殊案件
2. 参与判决的judges属于report中不可缺少的一部分，一般UKHL和UKSC的案件由五人作出裁决，所以在整篇文章的发言中我们需要区分每个法官对应的内容，在每个人发言的第一段前都会写明当前发言的法官。
3. 报告的最后都会标明committee对于该案件作出一个具体的裁判,表示支持或者驳回原告的诉讼请求，像这个案件里每一个法官都表明会allow the appeal, 而LORD SCOTT OF FOSCOTE 会再给出一系列论据之后说明，‘26. For these reasons I would allow the appeal and direct that the case be remitted to the County Court for trial.’
4. 而全文之中最长以及最重要的篇幅就是关于案件的judgement，其中会包含上诉理由，或the case stated by a lower court on an appeal by case stated。他会由一个主要法官或者是提供majority opinion的法官进行说明案件的背景，重要的事实以及引用过往可以依据的案例来代表所有人说明审判的过程和理由。除了向我们给的例子中体现的五个法官各自发言以外，随着发展出现了多位法官联合发言，即多人共同编撰了一些段落的judgement。这些不同的格式也为我们的分析带了一个难题，即哪一位法官的opinion在是majority opinion即超过半数以上法官同意的推理路线。

## National Westminster Bank Plc v. Amin and Another [2002] UKHL 9 (28th February, 2002)

#### HOUSE OF LORDS

Lord Nicholls of Birkenhead Lord Hoffmann Lord Hope of Craighead Lord Scott of Foscote Lord Rodger of Earlsferry

#### OPINIONS OF THE LORDS OF APPEAL FOR JUDGMENT

#### IN THE CAUSE

*NATIONAL WESTMINSTER BANK PLC (RESPONDENTS)*

*v.*

*AMIN (EXECUTRIX OF THE ESTATE OF MOHAMMED AMIN) AND ANOTHER (AP) (APPELLANT)*

**ON 28 FEBRUARY 2002**

**[2002] UKHL 9**

**LORD NICHOLLS OF BIRKENHEAD**

My Lords,

    1. I have had the advantage of reading in draft the speech of my noble and learned friend Lord Scott of Foscote. For the reasons he gives, and with which I agree, I would allow this appeal.

**LORD HOFFMANN**

My Lords,

    2. I have had the advantage of reading in draft the speech of my noble and learned friend Lord Scott of Foscote. I agree with it, and for the reasons which he has given I would allow the appeal and make the order that he proposes.

**LORD HOPE OF CRAIGHEAD**

My Lords,

    3. I have had the advantage of reading in draft the speech of my noble and learned friend Lord Scott of Foscote. I agree with, and for the reasons which he has given I would allow the appeal and make the order that he proposes.

**LORD SCOTT OF FOSCOTE**

My Lords,

5.This is a case which, in my opinion, must go to trial. It is a great pity that National Westminster Bank ('the bank'), respondent in this House, appealed against the decision to that effect reached both by Deputy District Judge Sanghera and by Judge Hindley QC. The bank appealed to the Court of Appeal and on 9 December 1998 Mummery and Clarke LJJ allowed the appeal. They held that the mortgagor, Mrs Amin, had no reasonable prospect of resisting the claim by the bank, as mortgagee, to possession of the mortgaged property, 82 Radnor Close. They applied principles set out in the judgment of Stuart-Smith LJ, giving the judgment of the Court of Appeal, in *Royal Bank of Scotland plc v Etridge (No 2)* [1998] 4 All ER 705. If the appeal in the present case had been heard after this House had dealt with the *Etridge (No 2)* appeal and the other conjoined appeals (see *Royal Bank of Scotland plc v Etridge (No 2)* [[2001] 3 WLR 1021](https://beta.bailii.org/cgi-bin/redirect.cgi?path=/uk/cases/UKHL/2001/44.html)), I think it likely that the appeal from Judge Hindley would have been dismissed.

…

26. For these reasons I would allow the appeal and direct that the case be remitted to the County Court for trial.

**LORD RODGER OF EARLSFERRY**

My Lords,

    27. I have read the speech of my noble and learned friend, Lord Scott of Foscote. I agree with it and for the reasons which he gives I too would allow the appeal and make the order that he proposes.

* **iv. Existing Tools**
* To gather information for our primary research aim, which was to develop a pipeline that could automatically produce case summaries in order to provide users personalized services, we found some existing technologies and resources that are useful for obtaining legal texts and summaries.
* Firstly we checked the BAILII website and there’s no summary option for the UKHL cases before 2009, and the UKSC document included a link jump to the summary provided by the Supreme Court official website.
* There are two types of summaries for each case. Taking [2021] UKSC 5 as an example, one is the case summary as shown in Figure 1. It only contains issues and facts which introduces the problems discussed in the entire report and the background of their appearance. However, it doesn’t contain the most important reasoning process or any views on behalf of the court. It can only help users to understand the case itself rather than the judgement.
* The website also provides a press summary option as shown in Figure 2. In addition to listing all the judges in the committee that heard the case and a more detailed background than Figure 1, it also provides the judgement and the corresponding reasons. We can find the reason section literally makes a summary based on each paragraph of the text and the references in square brackets are related to the paragraphs in the report. It is a formal summary given by professionals or the judges which can provide enough information to answer the questions in Figure 1. It can be used as the compared item to comparison with our generations which have a specific formatting such as adding paragraph numbers as evidence to tell users where the summary comes from. But works that requires human participation will also be time-consuming and have probability to have human mistakes. This is why we want to do more research on automated summary generation which not only help the court save time and labour costs, but also meet the personalised need of the final summary. We can change the academic level of the summary which means using different terms for uses with varies identities to avoid the general public not being able to understand professional terminology.
* 基于我们的主要研究任务-创建一个可以自动生成case摘要的pipeline为用户提供个性化服务，我们进行了一系列的调查发现了一些市场中现有的工具，这些工具在访问法律文本和摘要方面具有重要作用并被广泛应用。

1. 首先我们查看了BAILII网站，发现2009年以前的UKHL case没有summery的选项，而UKSC的document信息中包含了链接跳转到supreme court官网提供的summary
2. Supreme court官网针对每个case包含两种summary，以[2021] UKSC 5为例，一种是如图一的case summary，它仅包含issue和facts介绍了整个document讨论的问题和产生的背景，并不包含最重要的推理过程和任何代表法院的观点，只能帮助用户了解到事件本身而不是判决。该网站另外提供了press summary如图二， 他除了列举个审理此案的committee包含的所有法官和相比较图一更加详细的background以外，给出了judgement和对应的理由，从图中我们可以发现reasons部分基本根据正文的每一个段落做出了摘要，References in square brackets are to paragraphs in the judgment，它可以提供足够的信息去回答图一中的两个issue。它是一个正式的通过专业人士处理的summary，可以作为我们研究中用来比对的对象，我们可以学习他的排版方式比如添加段落的号码作为依据告诉用户这个结论从哪里来等等。但是需要人力参与的工作依旧会出现耗费时间并且有误差的情况，这就是为什么我们想要进行自动化摘要的研究，既可以帮助法院节约时间和人力成本，并且我们可以根据用户需求变换最后生成的摘要的学术程度，指针对不同身份的用户使用不同的用词，避免普通大众出现理解不了专业术语的情况。
3. A screenshot of a computer

   Description automatically generatedA screenshot of a computer

   Description automatically generated
4. LexisNexis is a powerful legal research tool, providing a wide range of legal text and summary information. Although LexisNexis Case Summaries provide a concise analysis method to help users quickly understand the issues in the case, they still have some limitations. LexisNexis provides a three-part analysis of the case, Procedural Posture, Overview and Outcome, but these summaries may be too brief and lack detailed legal and factual analysis. For some complex cases or situations that user require in-depth knowledge, this simplicity may limit the user's full understanding of the case. Besides this, LexisNexis summaries are only available on the LexisNexis platform and users must subscribe or purchase LexisNexis services to access them. This may limit the accessibility of these infomation, especially to researchers or students who do not subscribe to the LexisNexis service, and those general public who not usually use this .
5. In summary, although existing tools provide energetic and powerful function for users, they cannot meet the needs for large-scale legal text processing and personalized summary extraction. Therefore, our project aims to develop a new automatic summarization system that can handle large amounts of legal text more efficiently and extract personalized summary information based on user needs. By combining advanced natural language processing techniques with a deep understanding of the characteristics of legal texts, our projects will provide better tools and resources for legal research and practice.
6. LexisNexis被广泛认为是一种强大的法律研究工具，提供了广泛的法律文本和摘要信息。尽管LexisNexis Case Summaries提供了一种简洁的分析方法，帮助用户快速了解案件的问题，但它们仍然存在一些局限性：虽然LexisNexis提供了案件的三部分分析，Procedural Posture，Overview and Outcome: Describes the procedural disposition of the case.，但这些摘要可能过于简洁，缺乏详细的法律和事实分析。对于某些复杂的案件或需要深入了解的情况，这种简洁性可能会限制用户对案件的全面理解。LexisNexis摘要仅在LexisNexis平台上提供，用户必须订阅或购买LexisNexis服务才能访问。这可能限制了对这些摘要的可访问性，尤其是对于那些没有订阅LexisNexis服务的研究者或学生而言,而偶尔使用的大众更是对此没有需求。
7. 综上所述，虽然现有工具提供了一些法律文本和摘要信息，但它们存在一定的局限性，无法满足对大规模法律文本处理和个性化摘要提取的需求。因此，我们的项目旨在开发一种新的自动摘要系统，能够更有效地处理大量法律文本，并根据用户需求提取个性化的摘要信息。通过结合先进的自然语言处理技术和对法律文本特点的深入理解，我们的项目将为法律研究和实践提供更好的工具和资源。

**v. Summary**

* Recap of the legal background, highlighting key points relevant to the development of the automated summarization system.

1. **b. Technical Background**

Automatic summarization can be seen as an information compression process, which compresses one or more input documents into a short summary to provide users with concise text descriptions. This process inevitably involves information loss, but we must retain as much important information as possible. Through the introduction in section ‘legal background’, we found that the British law report has a specific format, the free expressing and wording style of different judges reflect the entire problem-solving process. Due to social and legal practice needs, legal workers and law students often need to quickly read the key information of judgment documents in order to get the idea. However, because the overall length of the document is usually longer or even nearly 100 pages, it is difficult to quickly understand the scattered key points of the judgment and inefficient to rely on manual case summaries by legal professionals. Therefore, automatic text summarization technology for legal judgment documents has become a practical and urgent practical need.

自动文摘的目的是通过对原文本进行压缩、提炼，为用户供简明扼要的文字描述。自动文摘可以看作是一个信息压缩过程，将输入的一篇或多篇文档压缩为简短的摘要，该过程不可避免有信息损失，但是我们要尽可能保留更多的重要信息。通过legal background介绍我们发现英国的law report框架结构具有特定格式, 里面不同法官自由的表达形式和用语方式体现了整个问题的解决过程。由于社会需求和法律实践的需要, 法律工作者、法学生往往需要快速阅读裁判文书的关键信息,以提炼总结裁判要旨.但因文书总体篇幅通常较长甚至近100页,快速理解散落的裁判要点非常困难,所以面对海量并快速增长的裁判文书,依靠法律专业人士人工编写的案例摘要将面临效率困境. 因此面向法律裁判文书的自动文本摘要技术成为了切实且紧迫的现实需求.

**i. Summarization**

There are currently two main methods of automatic summarization. Extractive summarization selects existing keywords and sentences from the original text to form a summary. It simply combines these sentences to create an article summary without modifying the sentences themselves. The technologies involved include keyword and key sentence extraction, and sentence semantic analysis, and the task of summarizing turns to find the most important sentences in the document, which is a sorting problem. This method uses ranking algorithms such as TextRank, which requires defining rules or feature sets to score and sort original sentences based on the features.

The other one is a abstractive method, which uses natural language processing to let machine analyse and understand the content of the article, and then uses natural language generation technology to automatically generate an summary that is different from the original sentences. Generally speaking, abstractive summarization can compress text more effectively than extraction, but it’s difficult to develop because they require the use of natural language generation technology, which itself is still in innovation. Generative technology requires the model to understand the semantics of the article and then summarize it, which is more similar to human practice. In 2014, when Bengio officially introduced the Sequence-to-Sequence model, which uses two recurrent neural networks to convert input text into vectors, and then convert the vectors into output sequences. This model was mainly used to complete machine translation tasks in the paper, and was later used in Google Translate, but it was also widely used in abstract generation tasks.

Here we continue Hachey and Grover's extractive summarization method for UKHL. Supervised automatic summarization extraction technology can be divided into machine learning-based methods and deep learning-based methods. Although the categories are different, they can be regarded as classification tasks or sequence annotation tasks when performing summary extraction. The automatic summarization method based on classification tasks uses a binary classifier to determine whether the sentences in the article belong to the summary. Common classifiers include SVM (support vector machine), DT (decision tree), CNN, RNN, etc. Supervised learning can better capture the rules contained in the text through the correspondence between the corpus and labels, but the classification-based method only independently examines the importance of the sentence and does not associate information such as contextual structure.

目前自动文摘的方法主要有两种, 抽取式文摘从原文中选取已存在的关键词和句子形成文摘。它将这些句子有机组合，得到一篇文章摘要，期间不修改句子本身。其中涉及到的技术有关键词、关键句抽取，句子语义分析，摘要的任务就变成了找到文档中最重要的几句话，是一个排序的问题。这种方法利用如 TextRank 这样的排序算法，**需要定义规则或特征集合**，根据特征对原文句子进行**打分**排序。

另一种是生成式的自动文摘方法，通过自然语言处理对文章的内容进行机器分析、理解，再使用自然语言生成技术，生成不同于文章中原有句子的自动文章摘要。一般来说，理解文摘可以比抽取更有效地压缩文本，但是可以做到这一点的自动文摘系统更难以开发，因为它们需要使用自然语言生成技术，而自然语言生成技术本身就是一个仍在不断发展的技术。生成式技术需要让模型理解文章语义后总结出摘要，更类似人类的做法，不过这种技术需要使用机器学习技术，长期以来并不成熟。转折点出现在 2014 年，Bengio正式引入了 Sequence-to-Sequence 模型，通过两个循环神经网络，分别把输入文本转化为向量，再把向量转成输出序列。这种模型在论文中主要用来完成机器翻译任务，并且后来被应用在谷歌翻译中，但后续在文摘生成任务中也产生了广泛的应用。

这里我们延续了Hachey and Grover对于UKHL的抽取式文摘方法，基于有监督的文摘自动提取技术可以分为基于机器学习的方法、基于深度学习的方法。尽管类别不同，但在进行摘要抽取时都可以视为分类任务或序列标注任务。基于分类任务的文摘自动提取方法使用一个二分类器确定文章中的句子是否属于摘要，常见的分类器有SVM(支持向量机)、DT(决策树)、CNN、RNN等。监督学习通过语料和标签的对应关系可以更好地捕捉文本中蕴含的规则，但基于分类的方法只是独立的考察了句子的重要程度并没有关联上下文结构等信息。基于序列标注任务的文摘自动提取方法采用序列标注模型，通过给每个句子打标签确定这个句子是否属于摘要。

**iii. Existing Corpus**

1. **HOLJ**:

Hachey and Grover published a new XML corpus of UKHL judgments, HOLJ, for use in their automatic summarisation system based on the approach of Teufel and Moens which consists of 188 judgments from the UKHL website between 2001 and 2003. The original HTML document is processed by a series of modules and converted into XML format, with annotation layers automatically added. There are two layers of manual annotation in the corpus,rhetorical role and the "relevance" of sentences based on whether they match sentences in the hand-written summary. In the current version of the corpus, there are 69 judgments have been annotated for rhetorical role, of which 47 have been annotated for relevance.

Legal judgments are more performative than research reports, and the judge's goal is to convince his professional and academic peers of the soundness of his arguments. Judgment is used to reflect its legal process, from legal interpretation to the final legalization decision. According to this feature, Hachey and Grover gave an overview of the interpretation of rhetorical notes as shown in Figure 1, which contains the tags FACT, PROCEEDINGS, BACKGROUND, FRAMING, DISPOSAL, TEXTUAL and OTHER. 'FACT' is the initial restatement of the judgment’s fact that make the appeal happens. When the judge restates the arguments and hearing details of the lower courts before UKHL hears the case, it will be marked as "PROCEEDINGS." ‘BACKGROUND’ refers to the authoritative precedents and legislation directly cited by the judge when considering the case. "FRAMING" is all aspects of the judge's chain of argument, and "DISPOSAL" refers to sentences in which the judge agrees or disagrees with a previous ruling. As shown in the frequency column in Table I, PROCEEDINGS, BACKGROUND, and FRAMING account for about 75% of the total sentences.

The corpus also contains manual summaries written by experts. Hachey and Grover used this content to compare with the original judgments and find corresponding sentences in the summaries as the basis for relevance judgments. 79% of the summary sentences have direct matches, 3% are direct connections, 9% are incomplete matches or connections, and 9% are summary sentences for which the corresponding sentence cannot be found.

These two parameters are also an important part of our research, because automatically generating summaries requires algorithms to complete these two tasks instead of human annotation, and their accuracy has become an important factor of how success the final pipeline is.

|  |  |  |
| --- | --- | --- |
| Label | Frequency | Description |
| FACT | 862 (8.5%) | A recounting of the events or circumstances which gave rise to legal proceedings e.g. On analysis the package was found to contain 152 mg of heroin at 100% purity |
| PROCEEDINGS | 2434 (24%) | A description of legal proceedings taken in the lower courts e.g. After hearing much evidence, Her Honour Judge Sander, sitting at Plymouth County Court, made findings of fact on 1 November 2000 |
| BACKGROUND | 2813 (27.5%) | A direct quotation or citation of source of law material e.g. Article 5 provides in paragraph 1 that a group of producers may apply for registration ... |
| FRAMING | 2309 (23%) | Part of the law lord’s argumentation e.g. In my opinion, however, the present case cannot be brought within the principle applied by the majority in.... |
| DISPOSAL | 935 (9%) | Either credits or discredits a claim or previous ruling e.g. I would allow the appeal and restore the order of the Divisional Court |
| TEXTUAL | 768 (7.5%) | A sentence which has to do with the structure of the document or with things unrelated to a case e.g. First, I should refer to the facts that have given rise to this litigation |
| OTHER | 48 (0.5%) | A sentence which does not fit any of the above categories e.g. Here, as a matter of legal policy, the position seems to me straightforward |

Hachey 和 Grover 基于the approach of Teufel and Moens发布了一个新的 XML corpus of UKHL judgments, HOLJ, for use in their automatic summarisation system. 这个语料库由来自上议院网站2001年至2003年的188个判决组成。原始的HTML文档经过一系列模块的处理，转换成了XML格式，自动添加了注释层。语料库中有两层手工注释。第一层是对句子进行修辞角色的手工注释，第二层是根据它们是否与手工撰写的摘要中的句子匹配来注释句子的“相关性”。在当前版本的语料库中，已对69个判决进行了修辞角色的注释，其中47个判决进行了相关性的注释。

法律判决比研究报告更强烈地具有执行性，法官的目标是说服他的专业同行和学术同行他的论点的合理性。判决则是用来体现它是通过合法的过程，从法律解释到最后做出合法化决定。根据这个特点，Hachey 和 Grover给出了修辞注释的解释概述如图一，其中包含标签FACT, PROCEEDINGS, BACKGROUND, FRAMING, DISPOSAL, TEXTUAL and OTHER.‘FACT’是判决的刚开始重述引起初始诉讼的事实和事件的部分。而当最高法院重述下级法院的审理论据和听证会细节到了上议院审理案件的时候，会被标为 “PROCEEDINGS”。‘background’则是在考虑案件时，法官直接引用的权威的先例和立法。 “FRAMING”是法官的论证链的所有方面，“DISPOSAL”指的是法官同意或不同意先前的裁决的句子，比如由于这是上诉法院，法官的实际决定，即允许或驳回上诉。 “TEXTUAL”修辞角色用于指示判决中的结构的句子，而“OTHER”类别用于不能放入注释方案中的句子。正如表I中的频率列所示，PROCEEDINGS、BACKGROUND和FRAMING约占句子总数的75％，而其他类别的句子频率较低。

语料库中还包含了专家们写的手动摘要，Hachey 和 Grover用这个内容与原文判决作比较，寻找摘要中对应的句子作为相关性判断的依据。其中79％的摘要句子具有直接匹配，3％是直接连接，9％是不完全匹配或连接，9％是找不到对应句子的摘要句子。这两个参数也是我们的研究中重要的一部分，因为自动生成摘需要让算法和机器代替专家完成这两个任务，他们的准确性也成了我们pipeline最后成功与否的重要因素。

**ASMO**:

ASMO is a corpus of 300 UKHL judgments created by J. Valvoda et al. in order to study majority opinion and agreement statements, and they asked three experts to identify the relevant AS and MO in each case. This includes 69 annotated data from HOLJ and 231 randomly selected cases out of 755 that the UK parliament website allows access to during 2001-2003. The final generation is a csv file in which each individual sentence becomes a line. Three experts individually annotated all 300 cases according to formal guidelines, identifying which cases contained sentences containing AS indicating full agreement and confirmation of which judges formed conclusive MOs based on these ASs. Agreement statements are those sentences that clarify the agreement between judges, for which a large number of formulaic phrases are proposed, and the judge expresses fully or partially agree by English words, as in the example in Figure 1.

ASMO是由J. Valvoda et al. 为了研究majority opinion 和agreement statements

创建的包含 300 个 UKHL 判决的语料库，并请三位专家识别每种情况下相关的 AS 和 MO。 这其中包含了HOLJ的69个有标注的数据，以及随机选取的2001-2003期间UK parliament website允许获取的755 个案中的231个。最后生成是每个单独句子成为一行的csv文件。三名专家依据正式的指导方针分别对所有 300 个案例进行了相关 AS 和 MO 注释，确定哪些案例中的句子包含AS，表示完全同意或确认以及哪些法官根据这些 AS 形成结论性的 MO。其中AS是那些明确了他们之间的协议的句子，为他们提出了大量公式化短语，法官从英语措辞中表达全部或部分协议，如图一中的例子。

**iv. The ASMO System**

• In the previous section we introduced the ASMO corpus, which contains agreement statements(AS) and majority opinions(MO) manually annotated by experts. At the same time, J. Valvoda et al. also introduced an AI system that automatically identifies these two elements with a performance comparable to humans.

They randomly divided the corpus into three subsets containing 100 cases, which were used as objects for different automation parts to avoid overfitting the training data, respectively called AS-set, MO-set and AI-set while they are used to train and test the machine learning model for the AS annotating classification, develop and test the rules for MO identification and evaluate the performance of complete AI system respectively.

The AS recognition task is a text classification problem, by comparing three different algorithms, Vector Machine (SVM), Logistic Regression (LR) and Naive Bayes (NB), to classify the sentences of the corpus into acknowledge, fully agree or other. This model is also inspired by Hachey et al.’s research on rhetorical role, which refers to dividing a text into different parts or paragraphs, and matching each part with a corresponding descriptor to better understand its structure and content. and Palau et al.’s research on argumentation mining, which extracts arguments, evidence and logical relationships from the text.

Valvoda et al.’s research utilize traditional feature sets including unigrams, part-of-speech tags (POS), sentence length, sentence position, and named entities (NE) to better understand the linguistic features and structure of the text, cue phrases manually selected by annotators during the annotation process are also collected as prompt phrase features like appeal, dismiss and reasons etc. POS and unigrams features are normalized by term frequency-inverse document frequency (TFIDF), which is a commonly used text feature weighting method that can reduce the impact of common words on classification and highlight those that appear frequently in specific documents but are not present in the entire corpus. They finally found that using LR achieved the highest F-score of 0.943 and could successfully combine all feature vectors but length.

* 在上一部分我们介绍了ASMO这个语料库，其中包含了专家手动标注的agreement statement and Majority opinion，而J. Valvoda et al. 同时introduces an AI system that automatically identifies this AS and MO with a performance comparable to humans.

他们将语料库随机平均分为三个包含100个案例的子集，作为不同自动化部份的对象以避免过度拟合训练数据，分别称为AS集、MO集和AI。

AS集用于训练和测试我们的ML模型，用于AS分类的子任务。MO集用于开发和测试我们的规则，用于MO识别的子任务。AI集用于评估通过将我们的AS分类ML模型与我们的MO识别规则相结合获得的完整AI系统。

AS集中的AS识别任务一个文本分类问题，通过比较三种不同的算法，向量机（SVM）、逻辑回归（LR）和朴素贝叶斯（NB），将语料库的句子分类为承认、完全同意或其他。这个模型也借鉴了Hachey et al.’s research on修辞分区指即将文本分成不同的部分或段落，给每一部分匹配一个对应的描述词，以便更好地理解其结构和内容。以及Palau et al.’s research on argumentation mining，从文本中提取出论点、支持论点的证据以及论点之间的逻辑关系。

在他们的研究中，针对法律文本分类任务除了利用了传统的特征组合包括单词、词性标签（POS）、句子长度、句子位置和命名实体（NE）以更好地理解文本的语言特征和结构，还收集了有注释员在标注过程中手动选择的常见短语作为提示短语特征。例如，“由于这些原因”、“允许/驳回上诉”或“我有幸”等。POS和单词特征进行了词频-逆文档频率（TFIDF）标准化处理，是一种常用的文本特征加权方法，它可以减少常见词语对分类的影响，突出那些在特定文档中频繁出现但在整个语料库中不常见的词语，从而提高模型的分类性能。他们最后发现使用LR获得了最高的F-score of 0.943 并且可以成功结合所有的特征向量but length。

MO refers to the set of reasons agreed by more than half of the judges in the court, and this aim is a complex process for the text. Figure 1 is an relationship diagram given by Valvoda et al., simulating the logical relationship between five judges in different cases whose complexity makes the final classification difficult.

The example (i) shows the D is MO as 3/5 judges {B,C,E} fully agreed {D}’s reasoning. although D did show to agree with someone else but express his own opinion so it is actually 4/5. We cannot confirm whether A has the same view as the majority of judges, but MO is not affected by A. Also in example (ii), 3/5 of the judges {B,D,E} identified {A,C} as MO, they believe that both A's and B's views are necessary to fully justify the decision.

Example (iii) is almost the same as the previous example, but shows a case where MO cannot be determined from AS. There are three groups here, {B,E} votes for {A,C}, while C only votes for himself and D only votes for A, which means there is no majority view. When a judge agrees with the reasoning of two or more judges, there is an 'and' relationship rather than an 'or' relationship, and agreement with the reasoning of judges A and C is different from only one of both. This also shows that MO cannot be determined simply by finding nodes with three or more incident edges.

Another complex situation in example (iv) also demonstrates this. Here, most {A,B,C} agree with D who agree with E. However, since D's reasoning depends on E, both of them must be contained in the MO, hence {D,E}. Note that since we are only considering fully agreement here, {A,B,C} are considered to agree with E, because if they did not agree with E, then they would not agree with D.

After adding judgment rules to the system and applied it on MO-set, this method finds (human-identified) MO from (human-classified) AS with 89% accuracy

MO 是指法院中有超过半数法官理由共同认可的集合，而这个问题对于文本来说是一个复杂的处理过程，图一是Valvoda et al.给出的关系示例图，模拟了不同案例里五位法官之间可能的逻辑关系，其复杂程度为最终的判断带来了困难。

在示例（i）中， 3/5法官{B,C,E} 确定了 {D} 为MO，因为前者都完全同意后者的推理。而D没有同意法官表明他以来自己的观点所以实际上是4/5.有时一个法官会写一个独立的首席裁决书，所有其他法官都同意这也会形成类似于图上展示的关系。而这张图里我们无法确认A是否与多数人的观点相同，但是MO不受A影响。

在示例（ii）中， 3/5法官 {B,D,E} 确定了 {A,C} 为MO，但是A和C都不是选举他们为MO的多数部分，因为虽然每个人肯定同意自己的推理但不能假设他们彼此之间也同意。但是大多数仍然认为，A和B的观点都是必要的，这样才能充分地证明决定的合理性。

示例（iii）与上一个例子几乎相同，但显示了一个无法从AS中推断MO的案例。这里面存在三种派别，{B,E}投票给{A,C}，而C投票给C但不投票给A，D投票给A但不投票给C，这意味着没有多数观点。当一名法官同意两名或两名以上法官的推理时，这之间是‘and’关系而不是‘or‘，与A和C法官的推理一致不同于与A或C法官的推理一致。这也表明MO不能简单地通过找到具有三个或更多入射边的节点来确定。

示例（iv）也证明了这一点。在这里，大多数 {A,B,C} 都同意D，后者又同意E。但是，由于D的推理依赖于E，因此它们两者都必须包含在MO中，因此是 {D,E}。请注意，由于我们在这里只考虑完全一致，所以A、B和C被认为与E一致-因为如果他们没有与E一致，那么他们就不会同意D。这是另一个说明这项任务有多困难的例子。

在给系统添加了判断rules之后，Applied on MO-set, this method finds (human-identified) MO from (human-classified) AS with 89% accuracy

**v. The SUMO System**

The HOLJ corpus mentioned in the previous section contains two important features, namely rhetorical role and sentence relevance ranking. Hachey and Grover used supervised learning in the SUM system to automatically identify the rhetorical labels of sentences in the judgment and another classifier to identify whether the sentences are relevant to final summary. They assigned a ranking score to each sentence which then is used to customize the length of the summary for the user. However, the F scores of these two classifiers are not perfect, being 60.6% and 31.2% respectively, which leaves room for improvement.

Subsequently, Imansyah merged the two corpuses HOLJ and ASMO to create a csv format folder containing all feature sets for each sentences and named it the ‘UKHL’ corpus. In addition to copying the six feature sets used by Hachey and Grover,

two new features, rhetorical labels and ASMO, are also introduced to train the classifier, where ASMO refers to the Boolean value of whether a sentence is an agreement statement and outcome. At the same time, the author also incorporated rhetorical tags into the relevance classifier, improving the classifiers’ F-score to 69.4% and 40.1%.

Subsequently, the SUMO system developed by Amy Conroy used newest package and tools to reproduce the classifier using spacy and Blackstone, which provided the possibility for us to apply it to new cases outside the corpus. The SUM system only used the sequence modelling in the classifier of rhetorical tags, while Conroy attempted to implement it for the relevance classifier. Through the study of conditional random field CRF, the f score of the two classifiers was finally improved to 77.8% and 42.1% which is based on a probabilistic graphical model. It takes into account the relationships between adjacent tags in the input sequence and predicts the most likely value of each tag based on these relationships. In other words, CRF take the overall structure of the judgement into consideration instead of just the independent features of individual sentence.

SUMO collects manually written ICLR case summaries from official website, and generates the final summary that meets the standards by adjusting the proportion of each rhetorical tag. The ROUGE-1 F1-score results indicate that the summary produced by SUMO (48.9%) perform better than summaries produced using the SUM methodology (37.6%) as well as the baseline summary (41.9%).

在上一部分提到的HOLJ corpus中包含了两个重要的特征为修辞角色和句子排名，Hachey 和 Grover在SUM系统里通过supervised learning来识别判决中句子的修辞标签和一个分类器识别句子是否相关，然后为每个句子赋予排名分数

然后通过这个排名来为用户定制摘要的长度，但是这两个分类器的f score不是很完美，分别为60.6% 和 31.2%，有很大的提升空间。

随后，Imansyah 将HOLJ和ASMO 两个corpus合并起来做成了一个包含所有特征值的csv文件夹命名为‘UKHL’语料库。在这个基础上接着研究两个classifier的提升方法，除了复制 Hachey 和 Grover 使用的六个特征集以外，

还引入了两个新特征，修辞标签和ASMO来训练分类器，其中ASMO指的是每个句子是否是agreement statement和final result的boolean value。同时他也将修辞标签纳入相关性中分类器中，将分类器 F 分数提高到了69.4% and 40.1%。

随后SUMO system developed by Amy Conroy 使用model pipline and tools在此基础上，使用spacy和Blackstone复刻了分类器，为后续我们将其应用在语料库以外的新case提供了可能性。 SUM系统只在修辞标签的分类器中使用了序列标记分类模块，而Conroy尝试为相关性分类器实现序列建模，证明了他在识别每个句子的相关性的可靠性。 通过对条件随机场CRF的研究，最后将两个分类器的f score提升至77.8%和 42.1%。

条件随机场（CRF）是一种用于解决序列标注问题的机器学习模型。CRF模型基于概率图模型，它考虑了输入序列中相邻标记之间的关系，并根据这些关系来预测每个标记的最可能值。换句话说，CRF考虑了输入序列的整体结构，而不仅仅是单个标记的独立特征。

Top of Form

SUMO收集了网站上手动编写的 ICLR 案例摘要，通过调节各个修辞标签的比例生成最后符合标准的摘要，使用ROUGE评估，The ROUGE-1 F1-score results indicate that the summary produced by SUMO (48.9%) perform better than summaries produced using the SUM methodology (37.6%) as well as the baseline summary (41.9%).

**vi. Summary**

* Recap of the technical background, emphasizing the tools, corpora, and systems relevant to the development of the automated summarization system.

# Chapter 3: Project Execution

1. Integration of Previous Pipelines

SUMO and ASMO were solutions studied in 2020 and 2018 respectively. When we wanted to integrate these two independent pipelines, we discovered a problem that they used different versions of Python and related libraries. Since these versions are outdated and incompatible, there are some challenges in integrating them into a modern pipeline.

Compared with the SUM system mentioned in the background, SUMO adds the NER tag provided by Blackstone which developed by ICLR. In addition to the original basic tags location, date, person and organization, Amy added entities, case name, citation, instrument, provision, court and judge specifically targeted at the legal field. This method increases the f score of rhetorical classifiers by 4%, indicating that these labels have a positive impact on the system. However, Blackstone has not conducted any maintenance since 2020. In order to use this toolkit in the SUMO project, Python 3.6, spacy 2.1.8, and numpy1.19.5 were selected because this is the highest configuration they can support.

However, in the Python ecosystem, libraries are updated very quickly, and new versions often introduce improvements and optimizations while fixing known issues and bugs. From the release cycle of python, as shown in Figure 1, we can find that Python 3.6 has already reached the end-of-life stage, which means that this version is frozen and no longer being updated. It is critical to Upgrading previous versions of pipelines to the latest version of Python to ensure that pipelines can take advantage of the newest features and performance optimizations, which enhance system stability and reliability. Simultaneously, spaCy supports a variety of basic functions of natural language processing, including word segmentation, part-of-speech tagging, named entity recognition, etc. After version 3.0, a new transformer-based pipeline was added, which also adds extra information and possibility that we can use to train new models later. In order to provide convenience and avoid frequent changes to the version of the tool library, we choose to update each related library version to ensure that the tool library can keep pace with the latest technology trends and updates, thereby providing better support for future research without the need of extensive code refactoring.

When we used the SUMO configuration with the necessary packages for ASMO, we found that the two pipelines could run successfully separately, which means that the ASMO environment can be backwards compatible even its Python version was 3.8 when it was developed. When we upgraded python to 3.11, both of them had problems. While Blackstone 0.1.15 in SUMO requires spacy in 2.1.8 version which means this need to be replaced completely to meet our requirments. After the pandas upgrade in ASMO, the pandas.core.indexes.numeric component was deleted which caused the pickle file generated before need to reproduce and modified the structure.

------------------------------------------

SUMO和ASMO是分别在2020和2018年研究的方案，我们要想将这两个独立的管道整合在一起时发现了一个巨大的问题，即它们分别使用了不同版本的Python和相关库。由于这些版本已经过时且不兼容，因此将它们整合成一个现代化的管道面临一些挑战。

A graph with colorful bars

Description automatically generated with medium confidenceSUMO中相较于在background中提到的SUM系统添加了ICLR中Blackstone提供的NER标签，在原先的基础标签location, date, person and organisation labels之外添加了专门针对法律领域的entities，case name, citation, instrument, provision, court and judge. 它的使用将rhetorical classifiers提高了 4%，表明这些label对系统有正面的影响。但是 Blackstone在2020年之后就没有再进行任何的维护，在SUMO项目中为了使用这个工具包，选择了Python 3.6 和 spacy 2.1.8，numpy1.19.5，因为这是他们是能支持的最高配置。

但是，在Python的生态系统中，库的更新速度非常快，新版本通常会引入改进和优化，同时修复已知的问题和漏洞。从python的release cycle 如图一我们可以发现，python 3.6已经在end-of-life state，这意味着这个版本is frozen，后续不会再进行任何的更新。将旧版本的管道迁移到最新版本的Python至关重要，以确保管道能够利用最新的功能和性能优化，提高系统的稳定性和可靠性。

同时spaCy支持多种自然语言处理的基本功能，包括分词、词性标注、命名实体识别、名词短语等，而3.0版本之后添加了新的基于 transformer 的 pipeline，为我们后续训练新的模型也提供了更多的可能性。

为了提供便利性并避免频繁更改工具库的版本，我们选择尽量将各个相关库版本更新，以确保工具库能够与最新的技术趋势和更新保持同步，从而为未来的研究提供更好的支持，不需要大规模的重构代码。

当我们使用SUMO的配置加上ASMO所必需的包时，我们发现两个pipeline可以成功运行，这意味着ASMO的环境可以向下兼容即使它本身开发的时候使用的python是3.8。而当我们将python升级到3.11，两个独立pipeline据出现了问题，SUMO中的Blackstone 0.1.15 requires spacy in 2.1.8 version，ASMO中的pandas升级后删除了pandas.core.indexes.numeric组件，这意味这前者在这个项目里的的完全废除而后者需要作出修改。

1. Both the rhetorical classifier and relevance classifier in SUMO use the ASMO tags in the study done by Valvoda et al. as the feature set. In ASMO, 'mj' in corpus is recorded as the judge name who made the majority opinion, ‘outcome’ refers to whether the sentence contains a statement of outcome. In Imansyah's UKHL corpus, the number recorded under the 'outcome' column is how many of the three individual annotators marked this sentence as an "outcome self" indicating this sentence contains the judges outcome. Both values are transferred to Boolean type during training as ‘asmo’ feature set. In previous research, Amy and Imansyah directly used the output data recorded in ASMO, which means they did not dynamically apply the pipeline to text to return the required feature values, but instead used the corresponding document generated previously with a direct merge. The limitation of this is that the research can only be applied to the same corpus without the ability to process new data, because their outputs are independent. In this part, we will modify our pipeline to allow inputting a plain text data to successfully reach the final step of outputting summary, without the need for excessive operations or manual processing.

In ASMO, the author uses AI set to evaluate the system which combined the AS classification model with the MO identification rules. This means that the output we should get at the end should be similar to the AI-set format, which contains the following columns as table 1, these are what we need to extract from the text. The most important function in ASMO provides the algorithm for extracting relation and mj, while other parts need to be solved by us. Here we all use [2001] UKHL 2 as the test case, corresponding to case 14 in ASMO corpus.

‘Casename’ is clearly stated as we generally process only one case at a time, which means that the content of this column can be the same as the file name itself, forming the primary symbol for this file.

Next, we need to deal with the problem of splitting sentences in the document. The first method we explored is to use NLTK to process this text file, obtaining the following results for its respondent and appellant ['PHILLIPS (LIQUIDATOR OF A.J.', 'BEKHOR & COMPANY) AND ANOTHER \nBREWIN DOLPHIN BELL LAWRIE (FORMERLY BREWIN DOLPHIN & COMPANY LIMITED)' ], and the '.' in the person's name is also separated where PHILLIPS (LIQUIDATOR OF A.J. BEKHOR & COMPANY) AND ANOTHER should appears as a whole phrase. In addition, the '...' symbol is processed into three separate sentences, which is an effect we do not want at all, showing its high reliance on punctuation marks.

Then we tested Spacy’s rule-based sentence segmentation method. It provides a pre-trained English model called "en\_core\_web\_sm", which uses a set of rules to determine sentence boundaries, based on punctuation and other language-specific cues. In the result, there’s decision that combine lord name with 'My Lords' but not every time like ['[2001] UKHL 2\nLORD STEYN\nMy Lords,\n1.\t', 'LORD HUTTON\n', 'My Lords, \n2.\tI have had the advantage of reading in draft the speech prepared by my noble and learned friend Lord Scott of Foscote...', which means we need to add more processing to adapt to the specific structure of the judgment like split at ‘\n’ sign. But the problem mentioned in NLTK didn’t occur with spacy, so we chose it as the main Natural language processing analyzer in our project. Because once the size of dataset increases, we cannot take consideration of all possible special cases and handle them in advance.

Sentence number, ‘line’ attribute here, identifies the order of each sentence in the document, while the ‘pos’ attribute indicates the proportion of each sentence in the entire document. We calculate the "pos" value of each sentence as the line number of the current sentence divided by the total number of sentences in the entire article. This approach helps us better understand the relative position and distribution of sentences in the document.

The 'from' column stores the identity of the speaker, but the speaker's name is not mentioned during the speech or in the sentence. We need to look for specific structural patterns in the text. We noticed that before each judge began to speak, there would be a specific paragraph in format of "LORD [name] My Lords." We view this specific structure as a signature factor that identifies different judges. Therefore, we developed an algorithm that locates and identifies this specific pattern in the text, and considered it as the starting point of a new judge's speech and prepend it with an identifier " -- NEW JUDGE --" to remind the system that the following content belongs to the same judge. This identifier is also one of the main parameters in the ASMO model.We extracted the name of lord to store in the ‘from’ column.

‘to’ column is the focus of our research in this section because this is the object of AS and MJ. First, there are labels for each sentence of the 100 cases marked in the annotator in AI set and we try to use these data to train a model to output the judge name of ‘to’. There are 46,893 rows in this data set, of which ‘NAN’ accounts for 45,008 (95%) and ‘self’ accounts for 470 (1%). The sign of imbalance lead to inaccurate trained models. Even when we used RandomForestClassifier to obtain an average accuracy of 0.974, the confusion matrix shows that the accuracy of NAN is 99% while the labels of various other lord names are all less than 40%, which means that our model is biased towards pointing to frequent appeared labels. And in these two sentences 'I have had the advantage of reading in draft the speech to be delivered by my noble and learned friend Lord Scott of Foscote. I agree with the order which he is to propose and with the reasons which he will give.', we can find that it cannot be judged from a single sentence. The ‘he’ in the second sentence here refers to Lord Scott of Foscote, but we did not consider the context when training the model. Simply extracting a sentence cannot be specified, resulting in an incorrect label.

---------------------------------------------

SUMO中的rhetorical classifier 和 relevance classifier均使用到了Valvoda 等人所做的研究中的 ASMO 标签作为feature set。虽然在corpus里’mj’记录为the judge name who made the majority opinion, 而outcome是指句子是否包含结果声明。在Imansyah 的UKHL corpus中，‘outcome‘ column下记录的数字是how many of the three individual annotators marked this sentence as an "outcome self" indicating this sentence contains the judges outcome.  在训练时这两个值都被转化为Boolean type组成了asmo feature set。在之前的研究中，Amy和 Imansyah都是直接使用了J. Valvoda et al.在ASMO中的输出数据，即没有动态的使用pipeline去处理文件返回需要的特征值，而是将之前生成的对应文档进行直接合并。这样做的局限性在于研究只可以应用在同一个corpus而不具有处理新数据的能力，因为他们的输出是分别独立的。在这一部分我们将修改我们的pipeline，允许输入一个纯文本数据就顺利到达最后输出summary的步骤，不需要做过多的操作或者人为处理。

在ASMO中, 作者使用AI 集来评估通过AS 分类 ML 模型与 MO 识别规则相结合而获得的完整 AI 系统，这意味着当我们最终应该获得的输出应该与AI-set格式类似，其中包含了以下几个column,这些也就是我们需要从text文本中提取出来的，而ASMO中最重要的功能提供了提取relation 和mj的算法,而其它部分需要我们来解决。这里我们均使用[2001] UKHL 2作为测试案例，对应ASMO corpus中的case 14.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | case |  |  |  |  |  |  |  | | line | body | from | to | relation | pos | mj |
| The case number specific in the corpus | The ordering of sentences in the entire document | The text of the sentence | The speaker of this sentence | Another judge if this sentence is a n agreement statement | |  | | --- | | full agreement or acknowledgement, | | And whether the sentence state the | | outcome | | Position of sentence in document. Range in (0,1) | The group of judge names for Majority opinion |

首先casename特别容易解决，我们一次一般只处理一篇文章，这意味着这一列的内容可以与文件名本身一样，形成这个文件唯一的指向符号。

接下来要处理文章的拆分句子问题。我们要探讨的第一种方法是使用nltk来处理这个文本，获得了以下结果['PHILLIPS (LIQUIDATOR OF A.J.', 'BEKHOR & COMPANY) AND ANOTHER \nBREWIN DOLPHIN BELL LAWRIE (FORMERLY BREWIN DOLPHIN & COMPANY LIMITED)'], 而人名里的‘.’也被单独分割，PHILLIPS (LIQUIDATOR OF A.J. BEKHOR & COMPANY) AND ANOTHER本身是作为完整的respondent出现的。除此之外‘…’符号会被处理成三个单独的句子，这是我们完全不想要的效果，显现出了他对标点符号的极度依赖。

接着我们测试了spacy的基于规则的句子分割方法。它提供了一个名为”en\_core\_web\_sm”的预训练英语模型，其中包含了一个默认的句子分割器。这个模型使用一组规则来确定句子边界，基于标点符号和其他语言特定的提示。比如有的会把lord name和‘My Lords’合并但不是每一次，['[2001] UKHL 2\nLORD STEYN\nMy Lords,\n1.\t', 'LORD HUTTON\n', 'My Lords,\n2.\tI have had the advantage of reading in draft the speech prepared by my noble and learned friend Lord Scott of Foscote ...',这意味着需要添简单的处理来在适应judgement里特定的结构。但对于nltk中提到的问题spacy没有发生，所以我们选择他来作为主要的nlp分析器，因为一旦后面数据增加，我们无法考虑到所有可能会出现的特例提前做出处理。

‘line’ 即sentence number 标识每个句子在文档中的顺序，而pos属性表示每个句子在整个文档中的位置比例。我们将每个句子的“pos”值计算为当前句子的行号除以整篇文章的句子总数量。这种方式有助于我们更好地理解句子在文档中的相对位置和分布情况。

‘from’column 储存的是发言者的身份，但是发言途中不会提及发言者的名字，我们在文本中寻找特定的结构模式。我们注意到在每个法官开始正式发言之前，会出现一段特定的内容，通常为“LORD [name] My Lords”。我们将这种特定结构视为识别不同法官的标志性因素。因此，我们开发了一个算法，用于在文本中定位并标识这种特定结构，一旦识别到这种结构，我们就将其视为新的法官发言的起点，并在其前添加了标识符“-- NEW JUDGE --”，以提示系统接下来的内容都属于同一个法官。这个标识符在ASMO模型中也是主要参数之一。

‘to’column时我们的研究重点，因为这是AS和MJ的对象。首先AI中有annotator中标注的100个case的每一句话的标签，我们尝试使用这些数据训练一个模型用来输出‘to’的judge name。在这个数据集里共有46893行，而其中‘NAN’占45008（95%），self占470（1%），这意味着我们数据不均衡会导致训练出来的模型不准确。即使当我们使用RandomForestClassifier获得了0.974的average accuracy，通过confusion matrix可以发现，NAN的准确率99%而其他各种lord name 的标签均低于50%，这意味着我们的模型偏向于将标签指向频繁出现的。而在这两个句子中‘I have had the advantage of reading in draft the speech to be delivered by my noble and learned friend Lord Scott of Foscote. I agree with the order which he is to propose and with the reasons which he will give.’，我们可以发现无法从单独的句子里判断。第二句里的‘he’在这里指代Lord Scott of Foscote，但是我们训练模型的时候没有考虑上下文，单纯抽出来一句话无法指明，导致错误的label。

|  |
| --- |
| . |
| . |

1. 重新设计和整合SUMO和ASMO管道的功能，以创建一个统一的现代化管道。这可能涉及到合并和重构代码，以确保其能够正确地处理数据流和实现所需的功能。
2. 进行充分的测试和验证，以确保新的管道能够正确地运行，并且能够达到预期的性能和稳定性要求。
3. 通过这些步骤，我们成功地重新实现了SUMO和ASMO管道，并将它们整合成一个现代化的管道，以确保其能够与最新版本的Python和相关库兼容，并且能够有效地处理数据并实现所需的功能。
4. Integration of Previous Pipelines
   * Discuss the challenges posed by the deprecated versions of Python and incompatible libraries.
   * Explain the need to reimplement the pipelines to ensure compatibility with the latest versions of Python and associated libraries.
   * Describe the steps taken to integrate the functionalities of the ASMO and SUMO pipelines into a single modern pipeline.
5. Replacement of Blackstone Features
   * Discuss the unavailability of certain features previously computed by the Blackstone extension of SPACY due to its deprecation.
   * Present the alternative methods or tools used to compute these features in the new pipeline.
   * Explain any adjustments required in the pipeline to incorporate these alternative methods.
6. Retraining Role and Relevance Classifiers
   * Discuss the necessity of retraining the classifiers, if applicable, to ensure their effectiveness with the updated pipeline and data.
   * Describe the process of retraining the classifiers and any adjustments made to accommodate changes in the data or pipeline.
7. Input from BAILI
   * Explain the need to obtain UKSC cases from BAILI instead of the UK Parliament website.
   * Describe the steps taken to modify the pipeline to accept input from BAILI.
8. Adaptation for UKSC Cases
   * Detail the challenges in applying the old pipelines to UKSC cases with a new co-authorship format for speeches.
   * Explain the modifications made to the pipeline to accommodate the new style of UKSC judgments.
   * Mention any adjustments made to process the new style of press summaries.

This chapter is intended to describe what you did: the goal is to explain the main activity or activities, of any type, which constituted your work during the project. The content is highly topic-specific, but for many projects it will make sense to split the chapter into two sections: one will discuss the design of something (e.g., some hardware or software, or an algorithm, or experiment), including any rationale or decisions made, and the other will discuss how this design was realised via some form of implementation.

This is, of course, far from ideal for *many* project topics. Some situations which clearly require a different approach include:

* In a project where asymptotic analysis of some algorithm is the goal, there is no real “design and implementation” in a traditional sense even though the activity of analysis is clearly within the remit of this chapter.
* In a project where analysis of some results is as major, or a more major goal than the implementation that produced them, it might be sensible to merge this chapter with the next one: the main activity is such that discussion of the results cannot be viewed separately.

Note that it is common to include evidence of “best practice” project management (e.g., use of version control, choice of programming language and so on). Rather than simply a rote list, make sure any such content is useful and/or informative in some way: for example, if there was a decision to be made then explain the trade-offs and implications involved.

## Example Section

This is an example section. This is an example citation [1]. The following is auto-generated dummy text. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a, venenatis vitae, justo. Nullam dictum felis eu pede mollis pretium. Integer tincidunt. Cras dapibus. Vivamus elementum semper nisi. Aenean vulputate eleifend tellus. Aenean leo ligula, porttitor eu, consequat vitae, eleifend ac, enim. Aliquam lorem ante, dapibus in, viverra quis, feugiat a, tellus. Phasellus viverra nulla ut metus varius laoreet. Quisque rutrum. Aenean imperdiet. Etiam ultricies nisi vel augue. Curabitur ullamcorper ultricies nisi. Nam eget dui. Etiam rhoncus. Maecenas tempus, tellus eget condimentum rhoncus, sem quam semper libero, sit amet adipiscing sem neque sed ipsum. Nam quam nunc, blandit vel, luctus pulvinar, hendrerit id, lorem. Maecenas nec odio et ante tincidunt tempus. Donec vitae sapien ut libero venenatis faucibus. Nullam quis ante. Etiam sit amet orci eget eros faucibus tincidunt. Duis leo. Sed fringilla mauris sit amet nibh. Donec sodales sagittis magna. Sed consequat, leo eget bibendum sodales, augue velit cursus nunc,

Figure 1: This is an example figure caption

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a, venenatis vitae, justo. Nullam dictum felis eu pede mollis pretium. Integer tincidunt. Cras dapibus. Vivamus elementum semper nisi. Aenean vulputate eleifend tellus. Aenean leo ligula, porttitor eu, consequat vitae, eleifend ac, enim. Aliquam lorem ante, dapibus in, viverra quis, feugiat a, tellus. Phasellus viverra nulla ut metus varius laoreet. Quisque rutrum. Aenean imperdiet. Etiam ultricies nisi vel augue. Curabitur ullamcorper ultricies nisi. Nam eget dui. Etiam rhoncus. Maecenas tempus, tellus eget condimentum rhoncus, sem quam semper libero, sit amet adipiscing sem neque sed ipsum. Nam quam nunc, blandit vel, luctus pulvinar, hendrerit id, lorem. Maecenas nec odio et ante tincidunt tempus. Donec vitae sapien ut libero venenatis faucibus. Nullam quis ante. Etiam sit amet orci eget eros faucibus tincidunt. Duis leo. Sed fringilla mauris sit amet nibh. Donec sodales sagittis magna. Sed consequat, leo eget bibendum sodales, augue velit cursus nunc,

for ( i = 0; i < n; i++ ) {

t[ i ] = 0;

}

Listing 1: This is an example listing

### Example sub-section

This is an example sub-section. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a, venenatis vitae, justo. Nullam dictum felis eu pede mollis pretium. Integer tincidunt. Cras dapibus. Vivamus elementum semper nisi. Aenean vulputate eleifend tellus. Aenean leo ligula, porttitor eu, consequat vitae, eleifend ac, enim. Aliquam lorem ante, dapibus in, viverra quis, feugiat a, tellus. Phasellus viverra nulla ut metus varius laoreet. Quisque rutrum. Aenean imperdiet. Etiam ultricies nisi vel augue. Curabitur ullamcorper ultricies nisi. Nam eget dui. Etiam rhoncus. Maecenas tempus, tellus eget condimentum rhoncus, sem quam semper libero, sit amet adipiscing sem neque sed ipsum. Nam quam nunc, blandit vel, luctus pulvinar, hendrerit id, lorem. Maecenas nec odio et ante tincidunt tempus. Donec vitae sapien ut libero venenatis faucibus. Nullam quis ante. Etiam sit amet orci eget eros faucibus tincidunt. Duis leo. Sed fringilla mauris sit amet nibh. Donec sodales sagittis magna. Sed consequat, leo eget bibendum sodales, augue velit cursus nunc, Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a, venenatis vitae, justo. Nullam dictum felis eu pede mollis pretium. Integer tincidunt. Cras dapibus. Vivamus elementum semper nisi. Aenean vulputate eleifend tellus. Aenean leo ligula, porttitor eu, consequat vitae, eleifend ac, enim. Aliquam lorem ante, dapibus in, viverra quis, feugiat a, tellus. Phasellus viverra nulla ut metus varius laoreet. Quisque rutrum. Aenean imperdiet. Etiam ultricies nisi vel augue. Curabitur ullamcorper ultricies nisi. Nam eget dui. Etiam rhoncus. Maecenas tempus, tellus eget condimentum rhoncus, sem quam semper libero, sit amet adipiscing sem neque sed ipsum. Nam quam nunc, blandit vel, luctus pulvinar, hendrerit id, lorem. Maecenas nec odio et ante tincidunt tempus. Donec vitae sapien ut libero venenatis faucibus. Nullam quis ante. Etiam sit amet orci eget eros faucibus tincidunt. Duis leo. Sed fringilla mauris sit amet nibh. Donec sodales sagittis magna. Sed consequat, leo eget bibendum sodales, augue velit cursus nunc,

**for** *i = 0* **upto** *n* **do**

ti ← 0

**end**

Algorithm 1 Example algorithm

#### Example sub-sub-section

This is an example sub-sub-section. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a, venenatis vitae, justo. Nullam dictum felis eu pede mollis pretium. Integer tincidunt. Cras dapibus. Vivamus elementum semper nisi. Aenean vulputate eleifend tellus. Aenean leo ligula, porttitor eu, consequat vitae, eleifend ac, enim. Aliquam lorem ante, dapibus in, viverra quis, feugiat a, tellus. Phasellus viverra nulla ut metus varius laoreet. Quisque rutrum. Aenean imperdiet. Etiam ultricies nisi vel augue. Curabitur ullamcorper ultricies nisi. Nam eget dui. Etiam rhoncus. Maecenas tempus, tellus eget condimentum rhoncus, sem quam semper libero, sit amet adipiscing sem neque sed ipsum. Nam quam nunc, blandit vel, luctus pulvinar, hendrerit id, lorem. Maecenas nec odio et ante tincidunt tempus. Donec vitae sapien ut libero venenatis faucibus. Nullam quis ante. Etiam sit amet orci eget eros faucibus tincidunt. Duis leo. Sed fringilla mauris sit amet nibh. Donec sodales sagittis magna. Sed consequat, leo eget bibendum sodales, augue velit cursus nunc,

Table 1: Example table caption

|  |  |  |
| --- | --- | --- |
|  |  |  |
| Foo | Bar | Baz |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# Chapter 4: Critical Evaluation

**A topic-specific chapter**

This chapter is intended to evaluate what you did. The content is highly topic-specific, but for many projects will have flavours of the following:

1. functional testing, including analysis and explanation of failure cases,
2. behavioural testing, often including analysis of any results that draw some form of conclusion wrt. the aims and objectives, and
3. evaluation of options and decisions within the project, and/or a comparison with alternatives.

This chapter often acts to differentiate project quality: even if the work completed is of a high technical quality, critical yet objective evaluation and comparison of the outcomes is crucial. In essence, the reader wants to learn something, so the worst examples amount to simple statements of fact (e.g., “graph X shows the result is Y”); the best examples are analytical and exploratory (e.g., “graph X shows the result is Y, which means Z; this contradicts [1], which may be because I use a different assumption”). As such, both positive *and* negative outcomes are valid *if* presented in a suitable manner.

# Chapter 5: Conclusion

**A chapter**

The concluding chapter of a dissertation is often underutilised because it is too often left too close to the deadline: it is important to allocate enough attention to it. Ideally, the chapter will consist of three parts:

1. (Re)summarise the main contributions and achievements, in essence summing up the content.
2. Clearly state the current project status (e.g., “X is working, Y is not”) and evaluate what has been achieved with respect to the initial aims and objectives (e.g., “I completed aim X outlined previously, the evidence for this is within Chapter Y”). There is no problem including aims which were not completed, but it is important to evaluate and/or justify why this is the case.
3. Outline any open problems or future plans. Rather than treat this only as an exercise in what you *could* have done given more time, try to focus on any unexplored options or interesting outcomes (e.g., “my experiment for X gave counter-intuitive results, this could be because Y and would form an interesting area for further study” or “users found feature Z of my software difficult to use, which is obvious in hindsight but not during at design stage; to resolve this, I could clearly apply the technique of Smith [7]”).

# Bibliography

|  |  |
| --- | --- |
| [1] | R. S. Laramee, “Bob’s project guidelines: Writing a dissertation for a BSc. in computer science.,” *Innovation in Teaching and Learning in Information and Computer Sciences,* vol. 10, no. 1, pp. 43-45, 2011. |

# Appendix A: An Example Appendix

Content which is not central to, but may enhance the dissertation can be included in one or more appendices; examples include, but are not limited to

* lengthy mathematical proofs, numerical or graphical results which are summarised in the main body,
* sample or example calculations, and
* results of user studies or questionnaires.

Note that in line with most research conferences, the marking panel is not obliged to read such appendices.