**西安电子科技大学**

**信息存储与检索 课程实验报告**

**实验名称 上机实验报告二**

经济与管理 学院 2006011 班

成 绩

姓名 马浩毓 学号 20069100105

同作者

实验日期 2023 年 5 月 11日

|  |
| --- |
| 指导教师评语：  实验内容：  实验效果：  实验报告：  考勤情况：  其他情况：  指导教师：  年 月 日 |

**一、实验目的：**

* 熟悉CNKI和万方数据库的基本情况；
* 熟悉Springer期刊全文和WOS数据库信息检索系统的基本情况；
* 掌握信息检索策略构建以及检索结果的分析；
* 掌握提高查全率和查准率的方法。

**二、实验要求**

1.检索课题名称

2.课题分析：确定课题中包含的关键检索词以及检索词的相关词

3.检索工具选择（并简述检索工具）；检索语言；检索途径；主题词选择

4.构建检索策略（检索表达式、时间范围、文献范围等）

5.检索策略调整的过程以及调整前后的结果（5-10条记录）

实验报告正文统一用小四号宋体,行距固定值20磅。要求：步骤合理、内容正确、项目完整、格式规范。

**三、实验内容：**

1、数据库基本信息了解

查阅该数据库收录的文献类型、服务功能、检索方式、检索流程等

2、选择和分析课题

对所选主题概念词从用、代、属、分、参几个角度进行词间关系分析，给出相应的与检索有关的相关词、同义或近义词以及上、下位词

3、构建检索策略。

确定包括所需课题的文献类型、检索式、检索的时间和空间范围、语种或其他约束条件、检索途径

4、检索策略调整

根据检索结果不断地调整检索策略，以提高查全率与查准率

5、检索结果分析

对于该课题相关的检索结果进行分析（所用检索工具（数据库）的名称；数量和期限记录；选择的学科类别名称、类号（从大类到小类）记录；选择的主题词记录；检索方式（途径）；检索表达式；实施的检索策略（该策略下检索结果数）、修改后的策略（检索结果数）；检索策略的调整记录、二次检索过程；5-10条命中记录的题录信息（题名、作者、文献出处等），CNKI和WOS检索结果的导出和字段的理解。

**四、实验步骤：**(需要截图，尤其是检索式、所操作关键界面等都需要截图.)

**1、数据库基本信息了解**

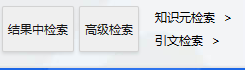
**知网（CNKI）**

知网（CNKI）是中国最大的综合性数字图书馆和学术搜索平台，其收录的文献类型包括：

1. 期刊论文：包括学术期刊、会议论文等。
2. 学位论文：包括硕士、博士学位论文等。
3. 会议文献：包括学术会议、技术会议等。
4. 专利文献：包括专利申请书、授权书、无效申请书等。
5. 标准文献：包括国家标准、行业标准、国际标准等。
6. 科技报告：包括科技研究报告、科技成果评价报告等。
7. 学术评论：包括学术期刊评论、学术会议评论等。
8. 政府出版物：包括政府出版的书籍、报告等。
9. 报纸文献：包括报纸、杂志等。
10. 其他文献：包括历史文献、综合文献等。



知网提供多种检索方式，包括：全文检索、关键词检索、摘要检索等、分类检索等。



检索流程一般包括以下几个步骤：

1. 输入检索词或关键词。
2. 选择检索方式（如全文检索、关键词检索、摘要检索等）。
3. 选择时间范围（如近一年、近三年、近五年等）。
4. 输入查询页码或者查询起止时间进行精确查询。
5. 显示检索结果，根据需要进行筛选。
6. 保存或下载检索结果。

通过以上步骤，可以快速准确地检索到所需文献信息。

**万方**

万方数据库是一个大型的文献信息服务平台，收录了大量的文献信息，包括期刊论文、会议论文、学位论文、专利文献、标准文献、政府出版物等。以下是万方数据库收录的文献类型、服务功能、检索方式和检索流程等方面的介绍：

文献类型：万方数据库收录的文献类型包括期刊论文、会议论文、学位论文、专利文献、标准文献、政府出版物等。



服务功能：万方数据库提供了多种服务功能，包括文献检索、全文阅读、摘要阅读、分类检索等。其中，文献检索可以按照主题、关键词、时间等多种方式进行查询；全文阅读可以浏览文献的全文内容；摘要阅读可以获取文献的摘要信息；分类检索可以按照主题、类别等方式进行检索。

检索方式：万方数据库的检索方式比较丰富，包括关键词检索、主题检索、时间检索、分类检索等。其中，关键词检索可以根据关键词进行精确查询；主题检索可以根据主题进行分类检索；时间检索可以按照时间进行模糊查询；分类检索可以按照主题、类别等进行分类检索。



检索流程：用户在使用万方数据库检索文献时，一般需要先输入关键词或主题，然后选择检索方式和时间范围，最后输入查询页码或者查询起止时间进行精确查询。查询结果出来后，用户可以根据需要进行筛选，保存或下载检索结果。

总之，万方数据库是一个比较全面和综合的文献信息服务平台，用户可以通过多种方式进行检索，获取所需的文献信息。同时，用户也需要根据实际需求进行精选和利用，提高检索效率和质量。

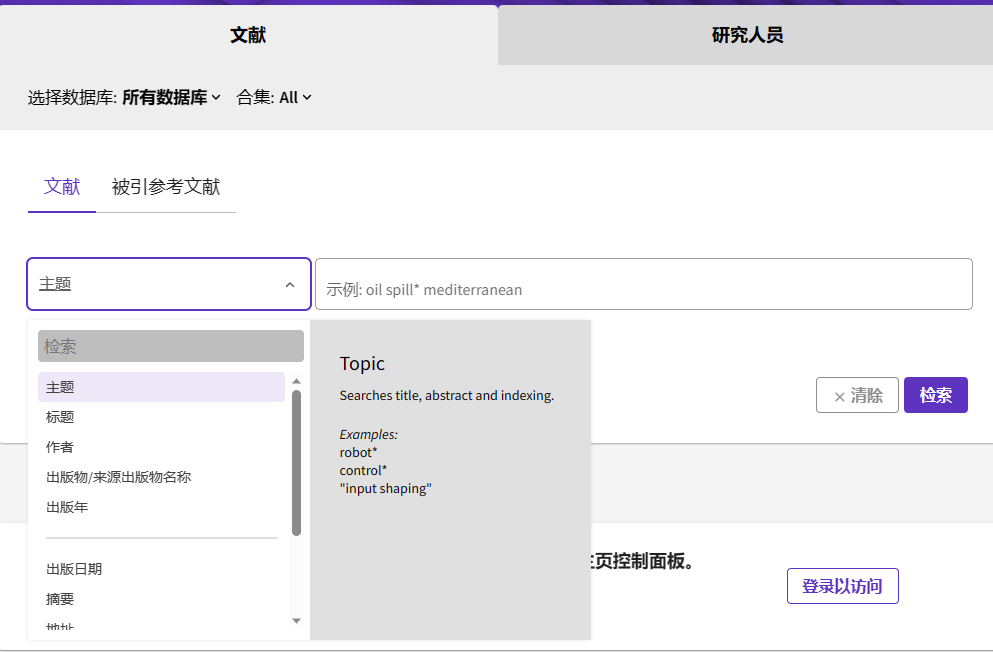
**WOS**

WOS（Web of Science）数据库是一个免费的、开放获取的期刊文献数据库，主要收录了全球学术期刊、会议文章等文献资料。以下是WOS数据库收录的文献类型、服务功能、检索方式、检索流程等方面的介绍：

文献类型：WOS数据库收录的文献类型包括期刊论文、会议论文、学位论文、专利文献、标准文献、政府出版物等。

服务功能：WOS数据库提供了多种服务功能，包括全文浏览、摘要阅读、分类检索、作者/题目字段检索等。其中，全文浏览可以浏览期刊论文的全文内容；摘要阅读可以获取文献的摘要信息；分类检索可以按照主题、作者/题目字段等方式进行检索；作者/题目字段检索可以根据作者/题目字段进行精确查询。

检索方式：WOS数据库的检索方式比较丰富，包括关键词检索、主题检索、时间检索、分类检索等。其中，关键词检索可以根据关键词进行精确查询；主题检索可以根据主题进行分类检索；时间检索可以按照时间进行模糊查询；分类检索可以按照主题、作者/题目字段等进行分类检索。

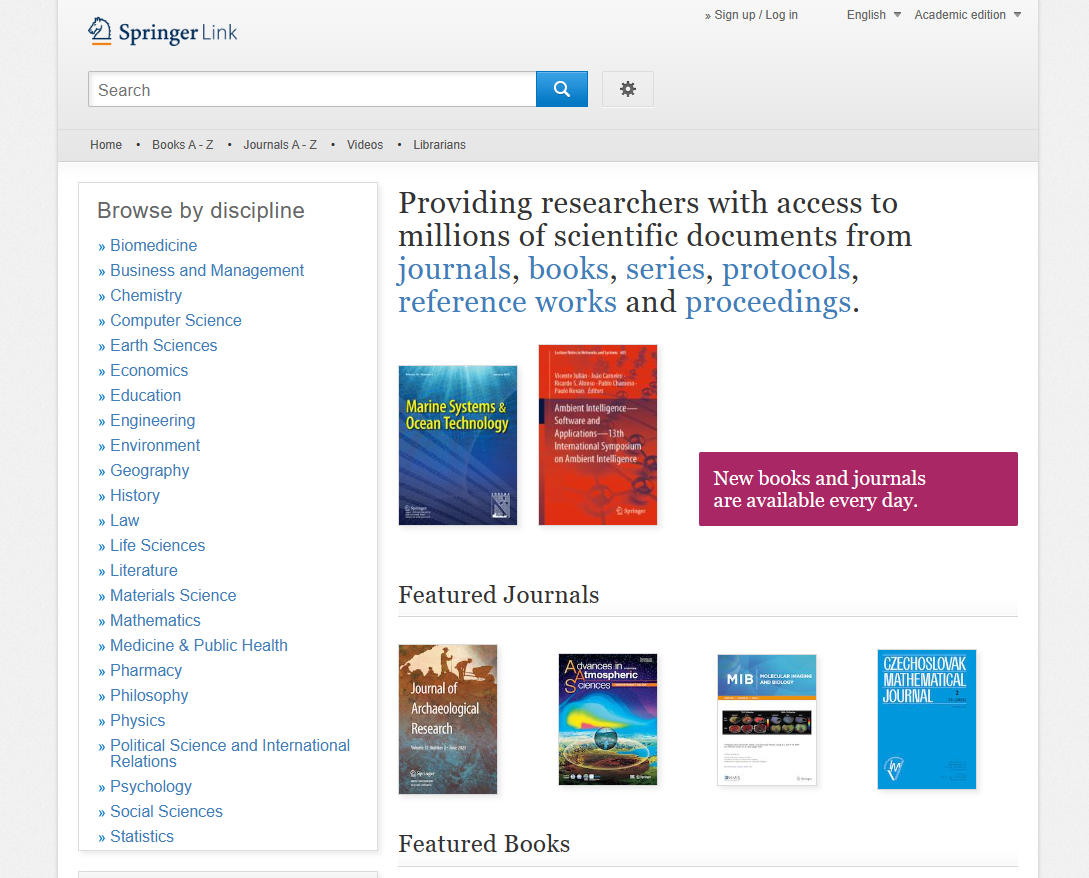


检索流程：用户在使用WOS数据库检索文献时，一般需要先输入关键词或主题，然后选择检索方式和时间范围，最后输入查询页码或者查询起止时间进行精确查询。查询结果出来后，用户可以根据需要进行筛选，保存或下载检索结果。

总之，WOS数据库是一个比较全面和开放获取的期刊文献数据库，用户可以通过多种方式进行检索，获取所需的文献信息。同时，用户也需要根据实际需求进行精选和利用，提高检索效率和质量。

**Springer**

Springer期刊是一个知识服务平台，提供了各种类型的文献，包括期刊、会议论文、学位论文、书籍等。其服务功能包括全文检索、摘要检索、引文分析、参考咨询等。



Springer期刊的全文检索功能可以帮助用户快速查找到文献全文，支持多种检索方式，包括标题、关键词、摘要等。检索流程一般包括输入文献标题或关键词，选择相应的检索方式，选择全文或摘要，进行查询和筛选。

除了全文检索外，Springer期刊还提供了摘要检索功能。用户可以在期刊中搜索文献并获取摘要，支持多种文献类型和格式，包括期刊、会议论文、书籍等。

Springer期刊的引文分析功能可以帮助用户了解文献之间的相关性和影响力。用户可以通过在期刊中搜索不同的作者、标题或关键词，获取他们之间的引用关系和引用频次，进而了解文献之间的相关性和影响力。

此外，Springer期刊还提供了参考咨询服务。用户可以通过在线咨询或电子邮件等方式向期刊提出问题或寻求帮助。期刊团队会根据用户提出的问题或需求，提供相应的解答或建议。

总之，Springer期刊提供了多种类型的文献，包括期刊、会议论文、学位论文、书籍等，并提供全文检索、摘要检索、引文分析、参考咨询等服务功能。用户可以通过在线咨询或电子邮件等方式向期刊提出问题或寻求帮助，获取相应的信息和建议。

**2、选择和分析课题**

这里我们选择主题词为：机器学习。

机器学习相关的词汇包括：学习、训练、算法、模型、监督学习、非监督学习、半监督学习、强化学习等。

机器学习的**同义词**包括：

1. 机器智能：指机器能够像人类一样理解、推理、学习和适应。

2. 计算机学习：指使用计算机程序来让计算机自主地学习新知识和技能。

3. 统计学习：指使用统计方法来构建机器学习模型和算法。

机器学习的**近义词**包括：

1. 深度学习：指使用多层神经网络来进行机器学习的方法。

2. 强化学习：指让计算机通过试错来学习最优策略的方法。

3. 监督学习：指使用带有标签的数据集来训练机器学习模型的方法。

机器学习的**上位词**是人工智能，指通过计算机模拟人类智能来解决问题的一门学科。

机器学习的**下位词**包括：

1. 监督学习：指使用带有标签的数据集来训练机器学习模型的方法。

2. 无监督学习：指使用没有标签的数据集来训练机器学习模型的方法。

3. 半监督学习：指使用带有少量标签的数据集和大量没有标签的数据集来训练机器学习模型的方法。

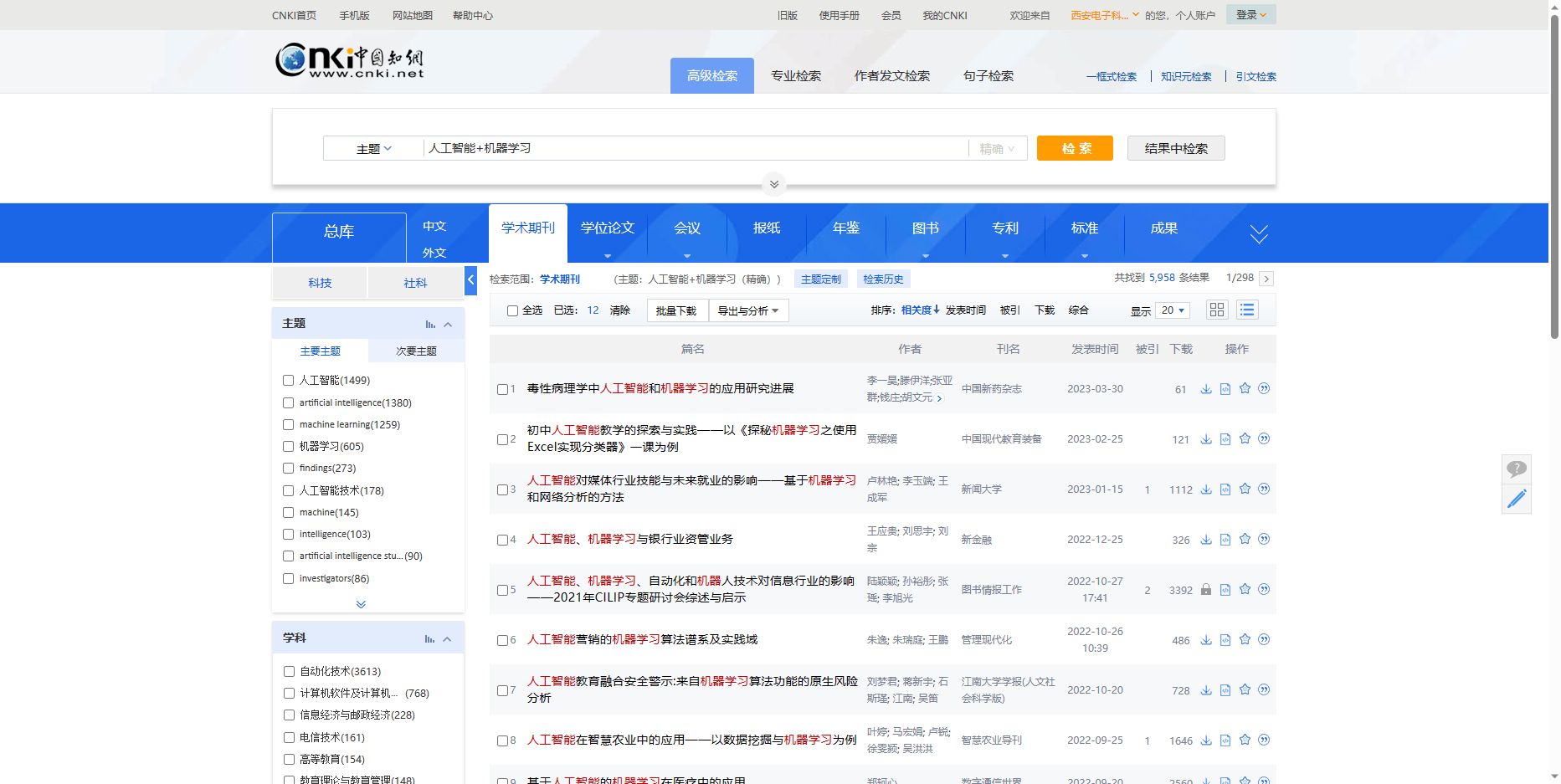
4. 强化学习：指让计算机通过试错来学习最优策略的方法。

**3、构建检索策略。**

检索平台选择CNKI，检索方式选择高级检索，选择文献类型为“学术期刊”，构建检索式为“人工智能+机器学习”，在主题中检索，检索时间范围不限，空间范围不限，选择中文数据库。



检索结果为：



共有5958条结果。

**4、检索策略调整**

由于机器学习为一个新兴领域，可以看出检索出的结果数量并不是很多，但还是包含许多诸如应用进展、大数据、人工智能时代等的文献，不是我想要的，因此需要进行检索策略调整。



第一次调整：添加检索式，检索主题不含有“大数据”或“人工智能时代”或“应用进展”的文献。



共4751条结果。

第二次调整：限制学科为“自动化技术”，得到如下结果：



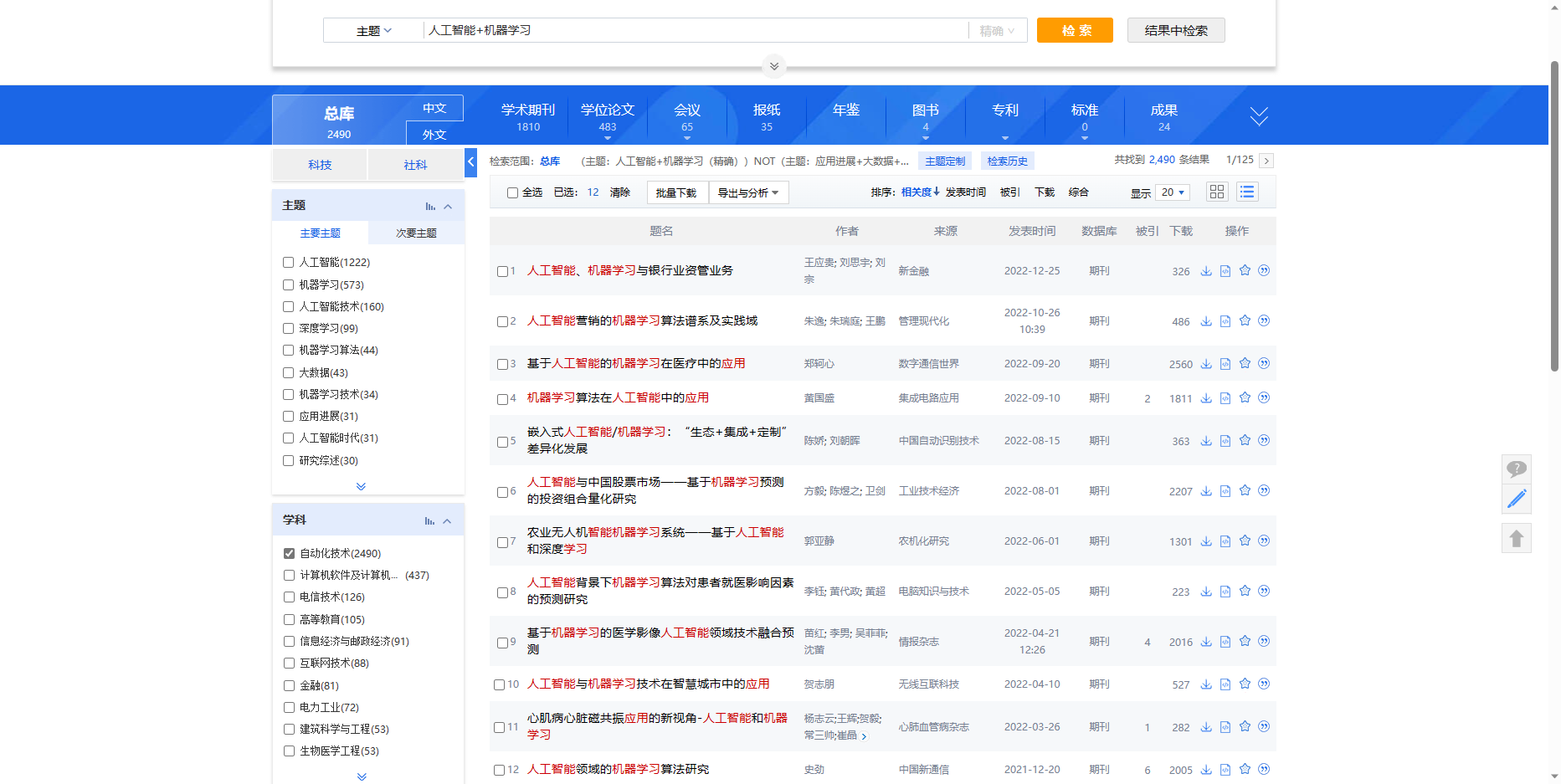
共2783条结果。

为了了解“机器学习”近期的研究，因此做第三次调整。

第三次调整：限制时间范围为2012—2022年。



检索结果如下：



共2490条结果，基本满足我的需求。

**5、检索结果分析**

此次检索，我所使用的数据库为CNKI，主题词为“机器学习”，检索方式为“高级检索”，检索表达式为“人工智能 + 机器学习”，检索范围为“学术期刊”；首次检索得到了5958条结果；共进行了3次检索策略调整，第一次调整为：修改检索表达式为(人工智能 + 机器学习) NOT (大数据 + 人工智能时代 + 应用进展)，得到了4751条结果；第二次调整为：限制学科为“生物学”，得到了2783条结果；第三次调整为：限制时间为“2013—2023年”，得到了2490条结果。

以下是部分命中结果的题录信息：





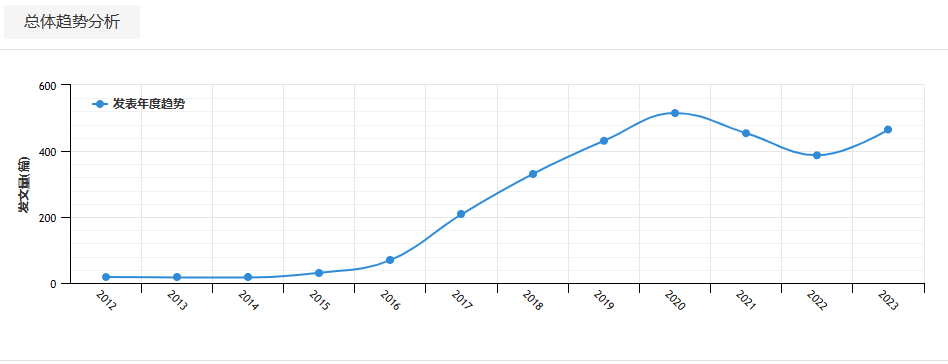






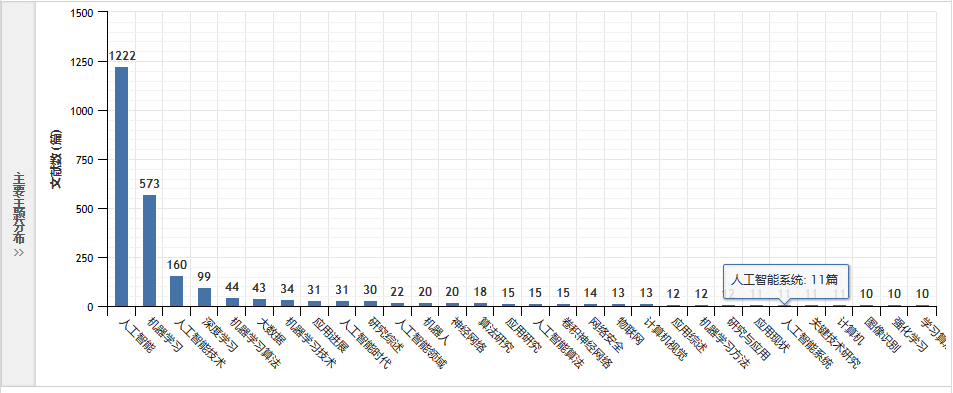
对检索结果进行导出分析，选择可视化分析，得到如下结果：

1. 总体趋势分析



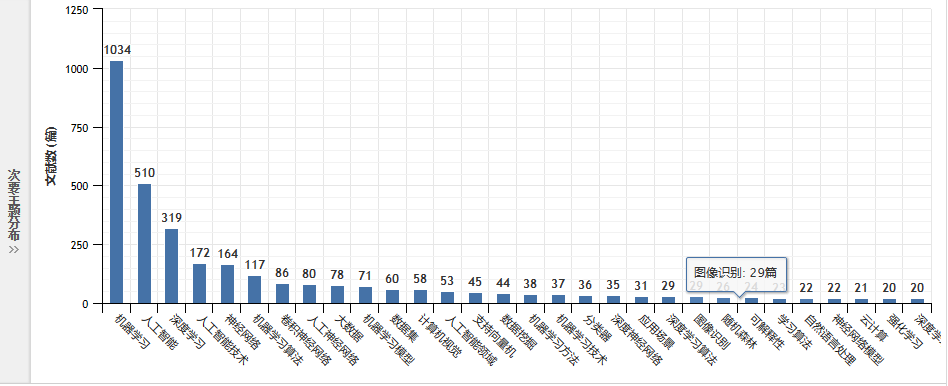
可以看到该领域近年来发文量与日俱增，是一个正在蓬勃发展的新兴学科。

1. 主要主题分布



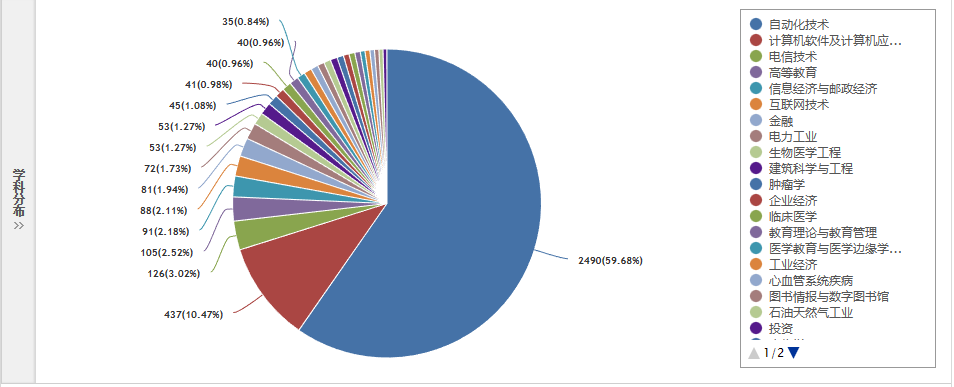
作为人工智能的一个分支领域，可以看到该关键词分布极端集中在“人工智能”和“机器学习”两个主题下。

1. 次要主题分布



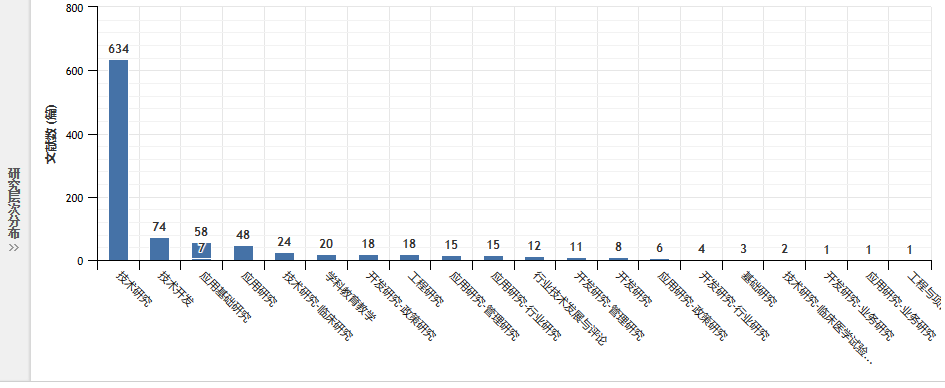
可以看到该新兴领域的次要主次与主要主题差别不大，“机器学习”的一些下位词领域也有所涉及。

1. 学科分布



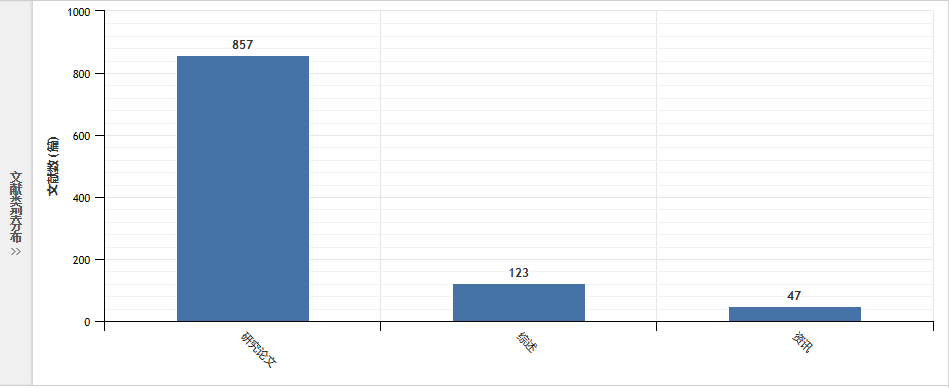
可以看到70%的成果都集中在计算机领域（自动化技术，计算机软件与应用）。

1. 研究层次分布



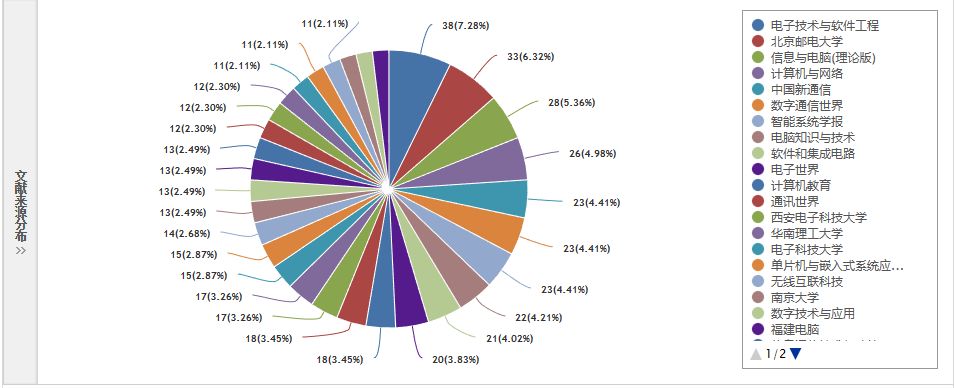
可以看到绝大部分研究集中在技术研究，因为机器学习领域主要由算法驱动，研究一个个不同的算法及其应用。

1. 文献类型分布



可以看到绝大部分文献为研究论文，因为该领域尚处于蓬勃发展阶段，理论和应用成果仍在增长，所以导致综述等文件较少。

1. 来源分析



可以看到文献来源相当多样，足见该领域的应用范围之广。

**五、实验结果：**

包括结果展示和结论；心得体会、尚存在问题等；进一步的思考等。

本次实验中，我先将不同文献检索工具进行了对比，然后我使用了CNKI数据库进行检索式构建，文献检索等工作，以机器学习为主题，我发现了该学科目前正处于蓬勃发展时期。研究型文献多，发表量成上升趋势。

通过此次实验，使我熟悉了CNKI、万方、Springer和WOS的基本情况；通过使用CNKI检索，使我掌握了信息检索策略的构建以及检索结果的分析，还可以通过对检索策略的不断调整优化，有效的提高查全率和查准率。