**利用协同过滤算法进行推荐**

本系统为工业从业者提供了技术问答，专业知识文献等方便而实用的功能，但是在繁多的信息中，找到自己想要看的资料可能需要花费大量的时间搜索，筛选，分辨……这时，一个稳定适用的推荐系统的出现显得尤为重要。

本系统主要采用基于物品的协同过滤算法，在用户的实际使用过程中，系统会记录下用户的各种行为，每一种行为在系统中都赋予了对应的权重。大量用户在一段时间的使用后，每份资料，包括技术问答，文章等等，都会形成一个用户评分矩阵，该矩阵为每个用户对这份资料的综合评分，而评分则是根据用户平时的行为进行加权计算得出。

最后，计算各份资料之间的余弦相似度，即可得到资料之间的相关度，接着便根据相关度进行推荐。

在这一过程中，如何计算权重决定了推荐系统是否实用。下面是本系统将用户的行为进行分类和加权的部分信息：

行为 权重

浏览问题 2

关注问题 3

回答问题 3

对问题的评论点赞 3

举报问题 2

浏览文章 2

收藏文章 4

点赞文章 3

点踩文章 2

评论文章 3

举报文章 2

以文章的行为举例，浏览文章作为最基本的用户行为，基于较低的权重2，当用户未对某篇文章有任何行为的时候，默认权重为1。而收藏文章则说明该文章对该用户的作用较大，因此基于较高的权重4。而点赞和评论文章说明用户一定完全并且较为认真地阅读过文章，这也能反应用户地兴趣点，因此给予权重3.而点踩和举报文章虽然能反映用户地认真阅读程度，但是这些负面行为同时也表明了这些文章无法对用户提供帮助，因此和基本浏览行为同权2。

**基于词频-逆文本频率指数的模糊搜索匹配系统**

当推荐的内容不能满足用户时，搜索是用户的不二选择。然而，搜索结果量大而杂，结果的前几项能高度符合用户的需求是必须的。由于本系统的特征是问答和文章交流，因此搜索结果也多为文献资料，采用词频-逆文本频率指数(tf-idf)能高效地表示两段文字地相关程度。

当用户在搜索框输入搜索内容时，若搜索内容为短语或较长的语句，则系统便会与数据库内积累的热门搜索项进行相似度计算，同时综合各个搜索项的热度给出最为相近的搜索词，以此来预测用户想要的输入。当然，tf-idf算法对短语之间的效果并不明显。当用户完成输入，正式进入搜索时，系统首先使用平常的模糊匹配关键字获取相关资料，然后计算资料内关键字的词频和逆文本频率指数，计算出tfidf值。最后按照tfidf值降序排列即为最终呈现给用户的搜索结果。

**采用卷积神经网络的智能分类**

本系统专注于特定几个方向的技术交流：减震降噪 传感器 新材料应用 绿色能源 射频技术 人工智能 流体机械 水处理技术 食品技术 制冷系统设计 燃烧技术 空气净化技术 机械结构设计 加热技术。当用户发布文章或问答的时候，智能地进行分类可以为用户的使用提供很大的便利，同时也减轻了管理员筛选分类资料的压力。因此本系统

使用CNN进行训练。网络结构为256核的卷积层，卷积核尺寸为5\*5，后跟最大池化层，最后为全连接层，并使用0.5的dropout率。在训练之前，需要对训练集进行预处理，第一步将所有训练集进行分词，并建立词汇表。词汇表中去除了常用并且不对文章意义产生影响的停用词

**Word2Vec词向量**

**OCR图像处理识别身份证信息**