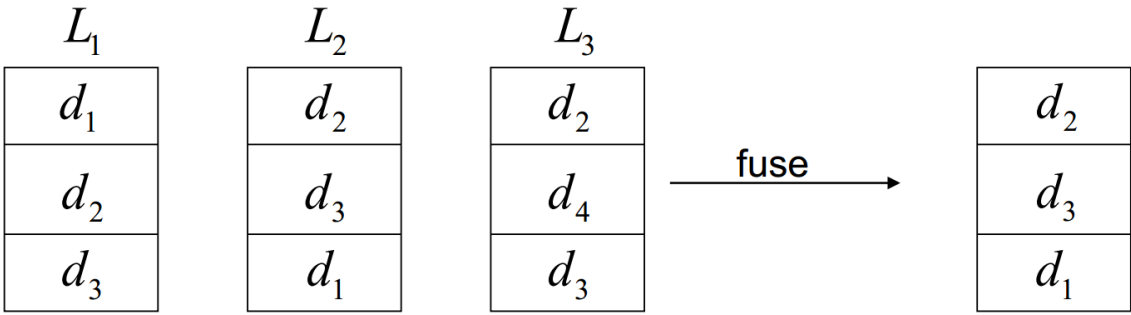


搜索结果融合

信息检索 信息融合 机器学习

搜索结果融合是将不同来源的结果进行合并融合，产出最后的搜索结果，包括: 聚合不同的文档(比如文本，图片，视频等)，不同的Query召回的结果和不同的排序函数得到的排序结果。可以应用到的场景: 搜索结果多样性，专家搜索，评估，查询表现预测，相关性反馈。



如图所示， L_1 ， L_2 ， L_3 分别代表不同的结果拉链，其中 d_1 ， d_2 ，...为拉链中的文档，多条拉链的结果经由融合过程后最后生成一条拉链即最终结果。

搜索结果融合
理论基础
计算上的社会选择理论
投票机制

理论基础

计算上的社会选择理论

将社会选择理论应用到计算问题上(比如将投票机制利用在排序聚合/融合上)，再利用计算框架分析和发明社会选择机制。

投票机制

4	3	2	2
Peter	Paul	Paul	James
Paul	James	Peter	Peter
James	Peter	James	Paul

Condorcet原则：Condorcet原则是指存在2个以上候选者时，若存在某个候选者能按过半数规则击败其他所有候选人则称该候选者为Condorcet候选人。

由上图可以看出一共有11次投票，其中4次Peter是在第一位的，因此Peter与Paul的对比结果为，在其中四次中Peter排名第一战胜Paul，在最后的两次中Peter排名第二也战胜了Paul，因此，Peter以6次战胜Paul超出了11次投票的半数，所以Peter胜出；Peter与James的对比结果中，其中四次Peter排名第一战胜James，在第三列中Peter又两次战胜James，因此同样以超出半数战胜了James，因此Peter成为Condorcet候选人。

Plurality原则：候选者在所有列表中排名第一的次数，次数最多的即为胜者。

可以看到Peter排名第一的次数为4，Paul排名第一的次数为5，James排名第一的次数为2，因此可以得出Paul为Plurality胜者。

Copeland原则：候选人两两比较，利用胜出的次数减去失败的次数，以此值决定胜者。

Peter对Paul胜出6次，失败5次，因此其值为1，Peter为胜者；Peter对James胜出6次，失败5次，因此取值为1，Peter为胜者。

Borda原则：一个候选人在一条拉链的分值是在一条拉链中排在它下面的候选人的个数。

Peter在四条拉链中得分为3分，3条拉链中得分为1分，在4条拉链中得分为2分，因此其最终得分为23分；Paul在四条来安中得分为2分，在五条拉链中得分为3分，在2条拉链中得分为1分，因此其最终得分为25分；James在四条拉链中得分为1分，在3条拉链中得分为2分，在2条拉链中得分为1分，在2条拉链中得分为3分，因此其最终得分为18分。因此胜出者是Paul。

