

Query理解 - 搜索引擎“更懂你”

原创 XG数据 WePlayData 2019-03-19

Query理解是对用户检索串分析、理解用户意图的过程，是一个标准的自然语言处理的任务。自然语言理解是一个非常复杂的过程，一方面体现在多词一义（同一种需求有若干种不同的表达）；另一方面体现在一词多义（苹果可能是水果，也可能是iphone手机）。现有自然语言理解的大量工作也都致力于解决多词一义和一词多义这两个问题。

下面以一个query为例介绍query理解的操作：

Query	蚊帐的失恋33天在线观看
Query纠错（qc）	文章的失恋33天在线观看
切词	文章 / 的 / 失恋 / 33天 / 在线 / 观看
Phrase	失恋33天
NER	文章（人名）；失恋33天（电影）
词权重	文章 0.45 / 的 0.04 / 失恋 0.82/ 33天 0.91 /在线 0.24/ 观看 0.22
一查非必留	的 在线 => 文章失恋33天换看
二查非必留	文章 的 在线 观看 => 失恋33天
同义词	在线观看 => 在线播放
时新性（qfrd）	0.2（时新性一般）
Query改写	放映失恋33天、文章的失恋三十三天
Query意图	影视 - 播放
Query商业价值	0.68（商业价值高）
topic model	娱乐，电影

从上文中看出实现搜索引擎“更懂你”，并不是很容易的事情，简单的一个query需要大量模块的分析和计算。query理解不仅帮助理解用户意图，同时对文档的召回和排序起着重要的基础作用。后续文章为依次具体介绍每个模块需要解决的问题和经典和前沿解决思路。

本文内容为星轨数据版权所有，未经许可许可不得任意转载复制，违者必究！