

今日看点 5min read

## 浅谈淘宝类目属性体系：商品搜索背后的逻辑架构



灿辉说搜索 2014/06/05



### II 摘要

淘宝拥有百万家商户和超过10亿的商品数，它如何让用户精准地找到想要的商品呢？其背后有着强大的技术支撑。

淘宝目前在线商品数超过 10 亿，如何精准的帮助用户找到他想要的商品呢？经过多年的探索，淘宝通过建立一套完整的类目属性体系，终于较好的解决了这一问题，今天就跟大家一起来谈谈淘宝的类目属性体系。

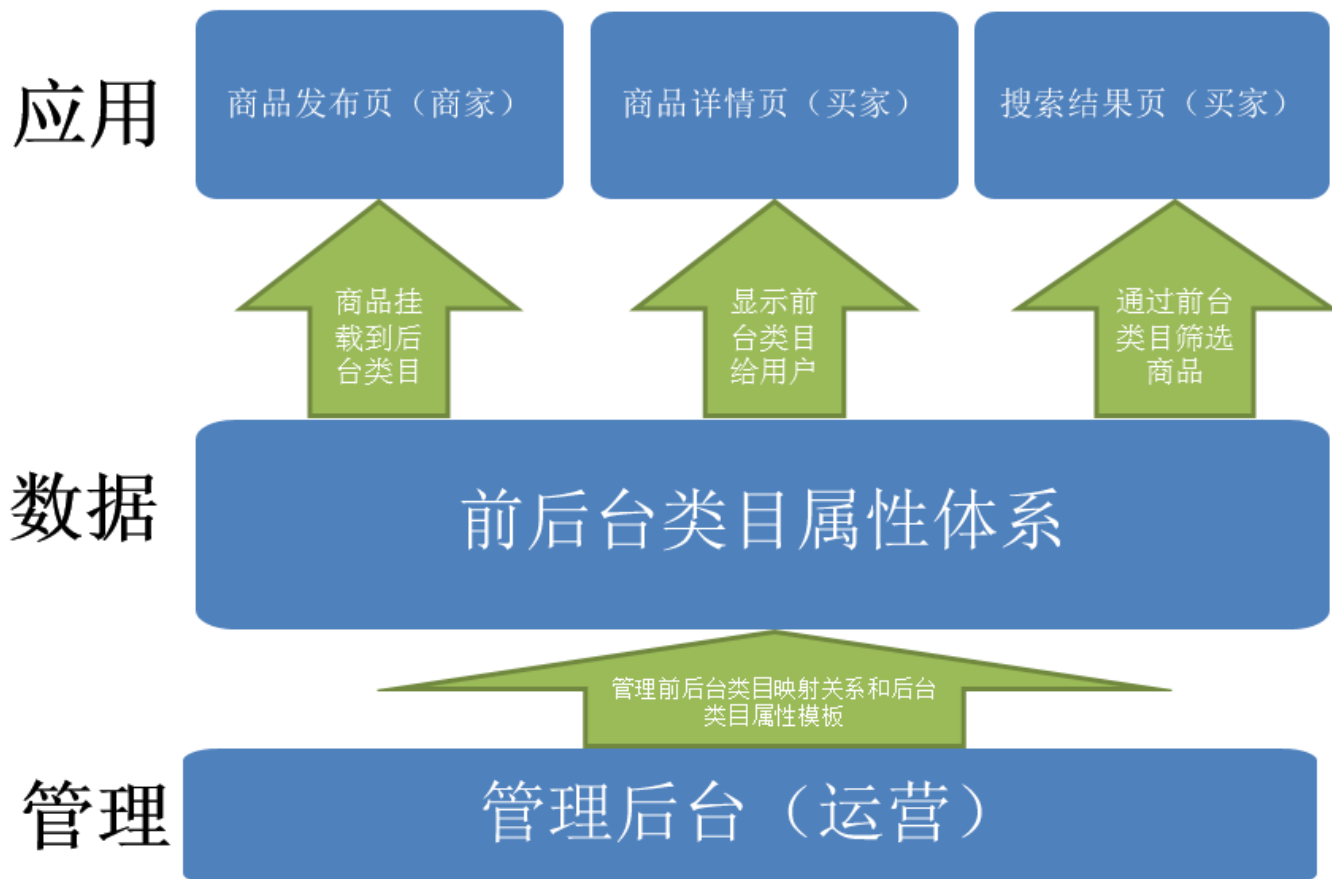
### 一点点历史和架构

2003 年淘宝刚上线时，商品量很少，没有分类。后来，商品量上百，开始有了对商品进行单级分类，有点类似于现在的一级行业类目。

了。从 06 年开始引入了属性，商家按照属性模板填写属性，用户可以按照属性筛选商品。

到了 08 年，开始将前后台类目分开，用户根据前台类目筛选商品，商家将商品挂到后台类目上，前后台类目树之间建立好映射。

今天的淘宝类目属性体系主要由后台类目树、前台类目树、挂载在后来叶子类目上的商品属性模板以及管理前后台类目之间映射关系的类目管理平台组成，整体架构如下：



从图中可以看出，淘宝类目属性体系是一个非常基础的数据服务，在商品发布页上商家选择后台类目上传商品信息，详情页上以面包屑的方式给用户显示商品所属的前台类目，在搜索结果页上让用户根据前台类目筛选商品。运营同学可以通过一个管理后台来管理前后台类目之间的映射关系以及后台类目的属性模板。

## 后台类目

后台类目面向商家，主要用于商品的分类和属性管理。商家上传商品时见到的就是后台类目，如下图：

淘宝网

宝贝 店铺

输入您想要的宝贝

Q 搜索



淘宝的后台类目树

后台类目有如下特点：

- 后台类目树中最重要的是叶子类目，也就是类目树上不能再往下分的类目,任何商品都必须挂载到后台叶子类目上。
- 叶子类目挂载属性模版，商家发布商品时选择好类目之后会根据属性模版，补充必填的商品属性信息，方可成功上传商品。
- 后台类目相对稳定，不能随便删除，叶子类目不能重复。

前台类目

前台分类面向用户，方便用户筛选查找商品，大部分时候用户见到的类目都是前台类目。下图就是淘宝搜索结果页上的淘宝前台类目树：



## 淘宝前台类目树的一部分

前台类目有如下特点：

用户购买时看到的类目是前台类目。

前台类目通过跟后台映射间接和商品关联。

前台类目由后台叶子类目+属性组成，一个前台类目可以包含多个后台类目，一个后台类目也可以映射给多个前台类目。

前台类目不直接挂载属性模版，前台类目对应的属性来自后台叶子类目的公共属性。

前台类目很灵活，可重叠，可删除。

### 为什么要将前后台类目分开呢？

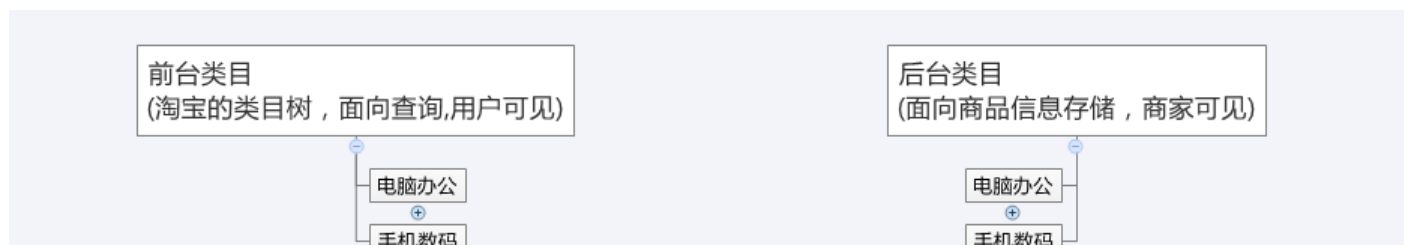
介绍完淘宝的类目属性体系，或许你会问为啥要将前台类目和后台类目分开呢？直接用一套类目体系不就行了吗？这里我就给大家解释一下原因，到了 08 年，淘宝商品日益增多，类目层级也越来越深，买家越来越难找到想要的商品。淘宝的小二们就开始不断调整类目属性，把类目树变浅，让商品更容易被买家找到。只有一套类目体系的情况下，小二每次调整，卖家也必须跟着调整。如果只是改一次，卖家也能勉强接受，但这是周期性的。比如，在夏天服装这个类目下可能是连衣裙、衬衫等夏装，到了冬天就会变成打底裤、羽绒服等冬装了，而且卖家也都知道这些调整是合理的，但是一年到头这么调整，意味着他一年到头都在编辑类目，可能还得专门养个把人在那里，一天到晚改类目，无形中卖家的运营成本也就提高了。

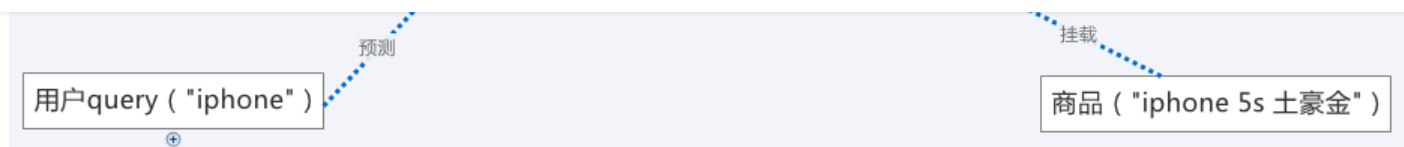
08 年的时候，有个淘宝的产品经理从线下零售巨头沃尔玛得到启发，他发现沃尔玛仓库里面的类目分区和货架区的类目分区是分离的，仓库里面的类目分区比较稳定，很少变化，而货架区的类目分区会根据活动和季节经常调整。类似的，淘宝也可以把原来的一套类目体系分为后台分类和前台分类，后台分类面向商家，用来挂载商品和属性模板，比较稳定很少变化；前台分类面向用户，主要方便用户查找商品，很灵活，可以经常调整。后台类目和前台类目之间通过映射联系起来，一个后台类目可以映射到多个前台类目，一个前台类目也可以包含多个后台类目。

从技术的角度来看的话，前台类目就是在后台类目的基础上建立了一个虚拟类目。

### 类目属性体系在搜索上的应用

先来看看类目属性体系是怎么帮助提升搜索的精准度的，下图是说明了用户 query 到商品搜索结果的全过程：





用户输入 query: iPhone, 淘宝类目属性体系预测到跟这个 query 最相关的前台类目是手机, 然后前台类目通过映射找到后台类目手机, 最终在这个类目下找到了“iPhone 5s 土豪金”这个商品。

有了类目属性体系, 用户搜索 iPhone, 淘宝就不会给出手机套这样的配件了, 因为它只会到手机类目下去找商品。

有了类目属性体系, 用户可以直接在搜索结果页按照类目和属性筛选商品, 更加高效的找到想要的商品。而淘宝搜索可以通过记录用户搜索 query 之后的类目属性筛选行为, 精准预测每一个 query 的类目和属性意图, 当用户再次搜索的时候, 可以根据 query 的类目属性意图, 在 query 文本相关性的基础上计算出商品的类目和属性相关性, 把包含 query, 且跟用户的类目和属性意图最相关的商品优先展示给用户。这也是百度这样的通用搜索引擎所做不到的, 因为它没有类目属性体系。

有了类目属性体系, 还可以做好搜索算法的垂直化优化。也就是在每一个细分类目下, 可以采用不同的商家权重和品牌权重, 甚至可以有不同的排序公式。比如用户选择手机时, 更多考虑各项参数和评测文章, 这个时候相应的排序权重因子可以加强, 而选择服装的时候, 一张漂亮的主图就无比重要了, 排序因子主图质量的权重就需要提升了。有了类目属性体系, 还可以做基于类目和属性的个性化搜索。搜索引擎通过用户的类目属性筛选日志, 可以挖掘出用户的一些特征, 比如性别, 意向价格区间, 偏好品牌, 偏好品类等, 从而可以对这个用户做个性化的展示。

## 类目属性体系无处不在

除了电商, 其实在各个需要帮助快速查找信息领域都可以见到类目属性体系的身影。比如: 教育行业里面的课程分类, 医疗行业的疾病和医院分类。当我们把用户查找的信息看做一个个实体的时候, 对实体分类以帮助快速定位查找就是一个非常通用的方法。从这个角度来讲, Google 的 Knowledge Graph 可以看做是将电商网站里面的商品类目属性体系推广到了整个世界中的万事万物, 它试图使用一套统一的类目属性网来囊括所有的实体, 或许这就是通用搜索引擎的类目属性体系吧。

题图来自 [Hybris Software](#)

作者: [@葛灿辉-搜索数据挖掘/搜索产品经理](#)

任职百度, 希望做一个有思想的搜索产品经理。

搜索

电商

淘宝

分享至

