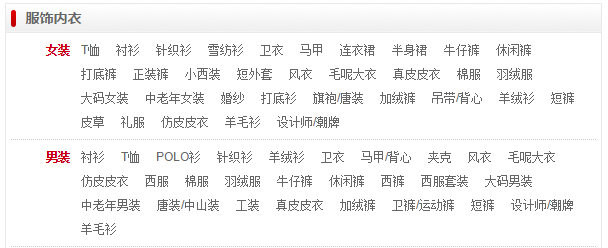
京东商品数据需求

1. 范围

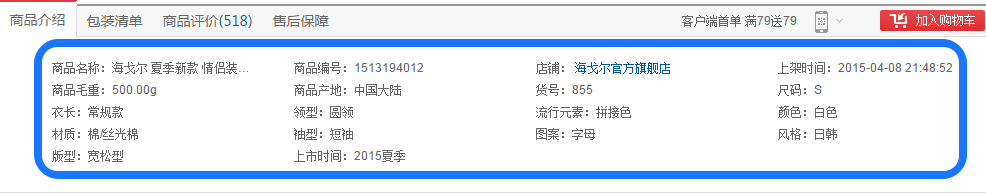
<http://www.jd.com/allSort.aspx>

女装、男装 下的所有子类



1. 具体需求（对每个商品页面）





* 1. 图片

前两张图（如存在）。

* 1. 文本信息

网页URL； √

页面上方的分类信息； √

商品名称； √

价格； √

商品介绍中的信息（除商品名称之外）； √

前两张图片的URL；

* 1. 多颜色商品

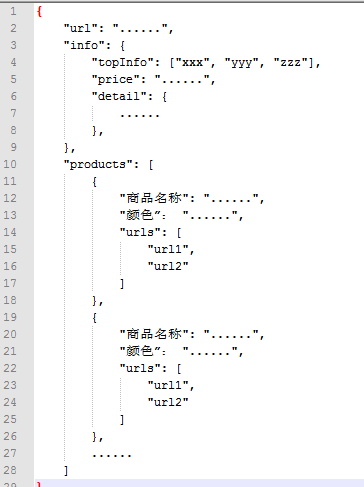
部分商品具有多个颜色，如下图



此种情况请**为每一种颜色分别**抓取前两张图片，及对应的产品名（选取不同颜色时品名会发生变化）

另外，对于多种尺码的情况不予额外考虑。

1. 建议数据格式
   1. 可采用JSON来存储每件商品的文本信息及网页与图片URL



* 1. 图片按照url中的文件名存在一个单独的folder中即可
  2. 因为图片较多，最好可以每个子类的文本和图片信息单独存放，即给“女T恤”，“男衬衫”等单独创建目录存放抓取的数据。