算法文字说明：

一.wordembedding：

1.导入全部招聘广告，jieba分词并生成总词语空间，应用word2vec算法，得到语料库中每一个词语的特征向量（本次向量维度取了10）

2.对应每一则招聘广告，加总其所有词语的词向量，得到每一则广告特征向量，进而得到总的招聘广告向量空间。

二：聚类算法：

改进型“k-means”算法：以每条语料间“余弦相似度”为相似度测度（原k-means算法以两点间欧氏距离为测度）

1.以k=20为例，其他情况同理。

①从之前得到的总的招聘广告向量空间中随机选取20则广告作为20个簇的质心。

②对于每一则招聘广告向量空间中的其他向量，分别计算他们与这20个质心的余弦相似度，将每一则广告划归到余弦相似度最大的簇下，意在说明它与该质心所代表的语义最为相近。

③在全部招聘广告向量空间中的向量都经过一次②过程后，重新计算20个簇的质心坐标。取簇内所有招聘广告向量对应分量的算术平均值作为新的质心坐标。

④重复过程②+③，直到簇的构成情况收敛于一个稳定状态，跳出算法。

⑤重复运行算法，调整k的值从10，11，12…….直到k=20。

⑤分析结果，人工判断聚类结果所反映的规律，观察每一个k取值的收敛结果。结果总是分类四类，分别对应着招聘广告技术要求的高、中、低、以及英文招聘广告。将招聘广告对于技术的要求的低、中、高与聚类结果中的相应簇进行对应，再次人工判断并将英文招聘广告放入对应的技术需求类别中，得以划分出最终的结果。