训练集数据说明：

train_author.json
*数据格式：*此文件中的数据组织成一个字典（dictionary, 记为 dic1），存储为 JSON 对象。 dic1 的键（key）是作者姓名。 dic1 的值（value）是表示同名作者集合的字典（记为 dic2）。 dic2 的键（key）是作者 ID， dic2 的值（value）是该作者的论文 ID 列表。

train_pub.json: 此文件包含 train_author.json 所有论文的元数据，数据存储为 JSON 对象；
*数据格式：*此文件的数据表示为一个字典（dictionary），其键（key）是论文 ID，其值是相应的论文信息。 每篇论文的数据格式如下：

域	类型	含义	举例
id	string	论文ID	53e9ab9eb7602d970354a97e
title	string	论文题目	Data mining: concepts and techniques
authors.name	string	作者姓名	Jiawei Han
author.org	string	作者单位	department of computer science university of illinois at urbana champaign
venue	string	会议/期刊	Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial
year	int	发表年份	2000
keywords	list of strings	关键词	["data mining", "structured data", "world wide web", "social network", "relational data"]
abstract	string	摘要	Our ability to generate...