1. **科学文献管理系统（难度★★★★）**

科研工作者的日常工作离不开查阅科学文献，并对其中的信息进行分析、筛选、挖掘和管理。请你为科研工作者设计一个管理系统，提高科学文献的管理效率。

* **目标用户**：科研工作者。
* **数据配置**：请通过以下方法下载数据文件dblp.xml.gz.

<http://dblp.uni-trier.de/xml/dblp.xml.gz>

将该数据文件解压后，其中包含一个dblp.xml文件。该文件由科学文献的记录序列组成，记录的格式如下所示：

<article mdate="2002-01-03" key="persons/Codd71a">

<author>E. F. Codd</author>

<title>Further Normalization of the Data Base Relational Model.</title>

<journal>IBM Research Report, San Jose, California</journal>

<volume>RJ909</volume>

<month>August</month>

<year>1971</year>

<cdrom>ibmTR/rj909.pdf</cdrom>

<ee>db/labs/ibm/RJ909.html</ee>

</article>

每个记录对应一篇文章，其中包含对作者，题名，发表杂志，卷号，出版时间等的详细说明。请基于该数据，设计能满足后述功能的文献管理系统。

**注：1）dblp.xml的大小超过1G，所以不要直接点击打开该文件。需要通过命令行命令’more’ 或者自行编程查看。**

**2） 编程过程中，不允许使用数据库系统。需要自己建立管理文献数据的数据结构。**

* **功能要求**：

F1. （必须实现） 基本搜索功能。输入作者名，能展示该作者发表的所有论文信息。输入完整的论文的题目，能展示该论文的其他相关信息

F2. （必须实现） 相关搜索。输入作者名，能展示于该作者有合作关系的其他所以作者。

F2. （必须实现） 作者统计功能。输出写文章最多的前100名作者。

F3. (扩展实现) 热点分析功能。分析每一年发表的文章中，题目所包含的单词中，出现频率排名前10的关键词。

F4. (扩展实现) 部分匹配搜索功能。给定若干个关键字，能快速搜索到题目中包含该关键字的文章信息

F5. (扩展实现) 聚团分析。作者之间的合作关系可以看成是一个图，每个作者对应一个顶点，任两个作者之间如果存在合作关系，则在两个顶点之间建立连边。这个图中的每一个完全子图我们称为一个聚团（所谓完全子图指的是该子图的任意顶点都和该子图的其他顶点有连边，完全子图的顶点个数称为该完全子图的阶数），请统计整个图中各阶完全子图的个数。

F6. (扩展实现) 可视化显示。通过图形化界面，展示作者之间合作关系图及其相关文章信息。