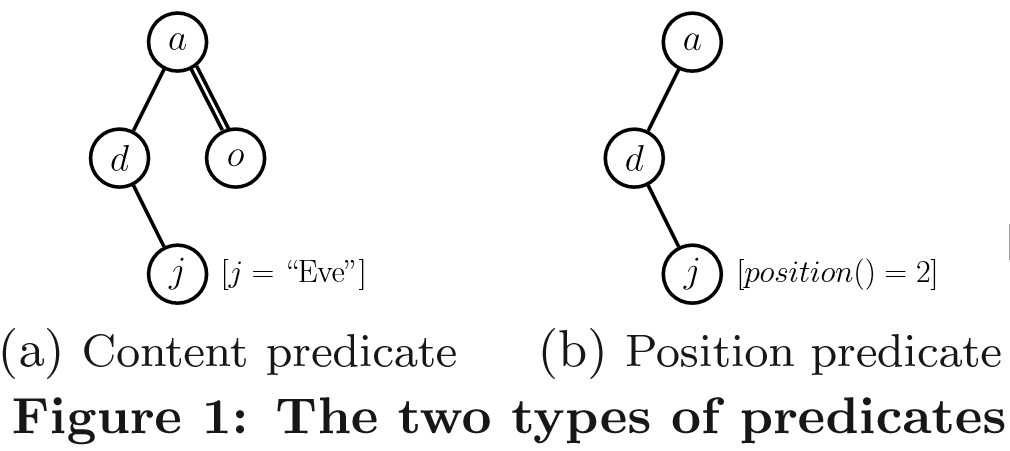
一、我们首先介绍一下这个方案的通信模型，这个模型总共有三个角色：数据用户Alice，数据所有者Bob和云服务器Charlie。首先，数据所有者将一组XML文档加密后，存储在云服务器上。然后，数据所有者可以对用户进行授权。被授权的用户就可以向云服务器发起查询，查询也是加密的。然后，服务器要通过加密的查询找到指定的加密文档，将它返回给用户。用户接收到查询结果后，就可以将其解密，获得所需的文档。

二、接下来我介绍一下这篇论文用到的一些定义

定义1（树模式）：树模式是一颗树，另外它需要满足两个条件：1）它的每个节点都要有名称；2）它的每条边应该是表示父子（PC）关系的单边或者是表示祖先（AD）关系的双边。

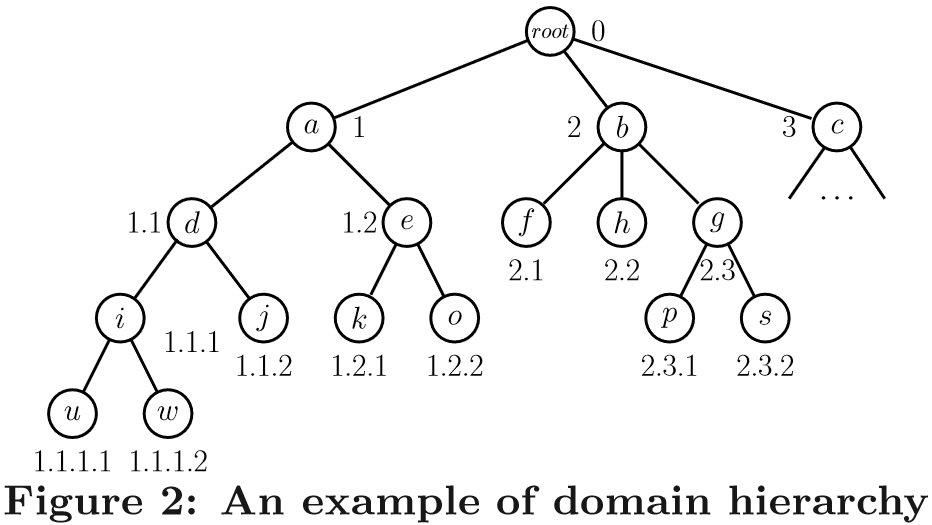
定义2（树模式查询）：树模式查询是一个查询，它可以表示为一对数据（Qt，Qc），其中Qt是树模式，Qc是在Qt的节点上定义的谓词的集合。也就是说，一个树模式加上一个谓词集合就形成了一个树模式查询，用户向云服务器发起的查询就是树模式查询TPQ。

而谓词集合中的谓词分为两种。第一种是内容谓词，可以表示为[x op α]，比如[x < 5]，意味着x节点上的值 < 5。op∈{>，≥，<，≤，=}，α是一个值。第二种谓词是位置谓词，可以表示为[position() = m]。它表示当前节点是其父节点的第m个子节点。



1. 解决方案

介绍完定义后，接下来我要介绍具体的解决方案。

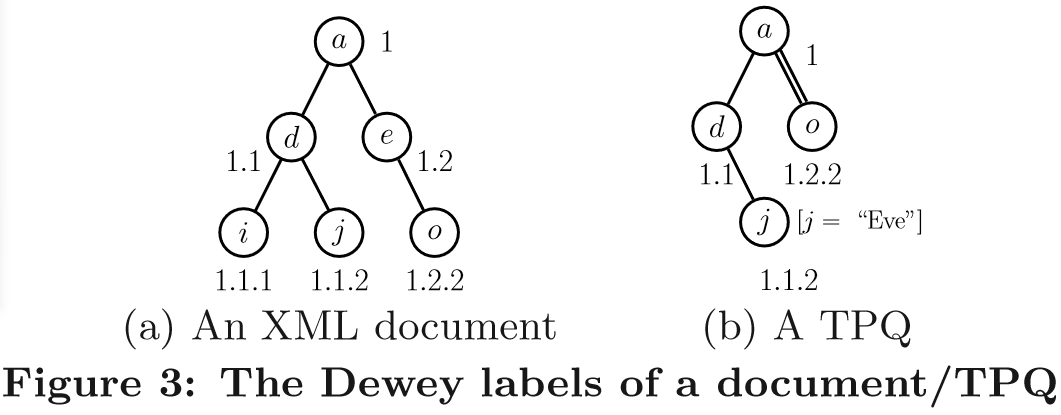


我们假设在这个系统中总共有3个XML数据库，分别表示为数据库a,b,c，这3个数据库都有各自的文档类型定义（DTD）【文档类型[定义](https://baike.baidu.com/item/%E5%AE%9A%E4%B9%89" \t "https://baike.baidu.com/item/%E6%96%87%E6%A1%A3%E7%B1%BB%E5%9E%8B%E5%AE%9A%E4%B9%89/_blank)（DTD，Document Type Definition）是一种特殊文档，它规定、约束符合标准通用标示语言（SGML）或SGML子集可扩展标示语言（XML）规则的定义和陈述】。而每个数据库中的XML文档都是根据其文档类型定义（DTD）来构建的，也就是说，如果我们把数据库的文档类型定义看成一棵树，那么这个数据库中的XML树就都是这颗DTD树的子集。然后，我们在这3个DTD树上添加一个根节点，构建成一颗大树，我们把这棵树称为域层次结构。这样的话，这个系统中的所有XML文档都可以看成是这个域层次结构的子集，而这篇论文的作者把这种现象称为嵌入，XML文档是域层次结构的嵌入。

我们的第一步是生成域层次结构，第二步就是**用Dewey 标记方案[25]来标记这棵域层次结构树。** 我们对树中的每个节点都进行标记，然后我们对这些标记进行排序。

顺序：1，1.1，1.1.1，1.1.1.1，1.1.1.2，1.1.2，1.2，1.2.1，1.2.2，2，......

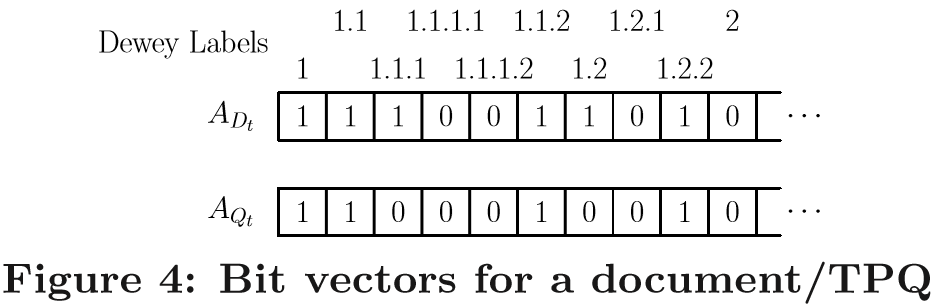
接下来，我们参照域层次结构中的节点的标记对数据库中的XML文档以及用户发起的查询TPQ进行dewey标记。我们假设有一颗XML树的标记是这样的。而因为域层次结构中可能有多个名称相同的节点，所以查询树可能有多种标记方案。我们假设有一个TPQ的其中一种标记是这样的。



然后，我们要将这两棵树编码为位向量。这个是上面提到的域层次结构中节点标记的排序。下面我们用ADt表示文档的位向量，用AQt表示查询的位向量。如果XML树中的某个节点在域层次结构中存在对应的节点，那么就将位向量中节点对应的位置设为1，否则就设为0，这样全部设置好后就将XML文档编码成了位向量，它可以表示XML文档的结构，所以我们把它称为结构向量。同样的，我们生成TPQ树的结构向量。

接下来，我们就可以检验文档和查询的结构是否匹配了，我们只要求出这两个向量的向量积，然后比较AQt中1的个数和向量积是否相等，就可以判断结构是否匹配了。

数组：1，1.1，1.1.1，1.1.1.1，1.1.1.2，1.1.2，1.2，1.2.1，1.2.2，2，......

 接下来，我们要进行内容匹配。我们首先要生成XML文档的文本向量。生成过程就是，如果XML树中的某个节点具有值，而且它在域层次结构中存在对应的节点，那么我们就将这个节点的值保存到文本向量ADc中对应的位置。

接下来，我们要对XML文档的文本向量ADc和TPQ中的谓词集合Qc进行计算，用来判断查询和文档的内容是否匹配。论文的作者首先使用另一篇论文中描述的一种叫做ASPE的加密方案，来加密ADc和Qc，然后将它们作为参数用公式进行运算，最后根据运算结果来判断文档D是否满足查询TPQ的谓词条件。因为公式比较复杂，这里就不详细介绍了。

这个方案检验查询和文档是否匹配的大概过程就是这样了，论文后面介绍的就是怎样提高这个方案的安全性和性能，在这里我也不详细介绍了。