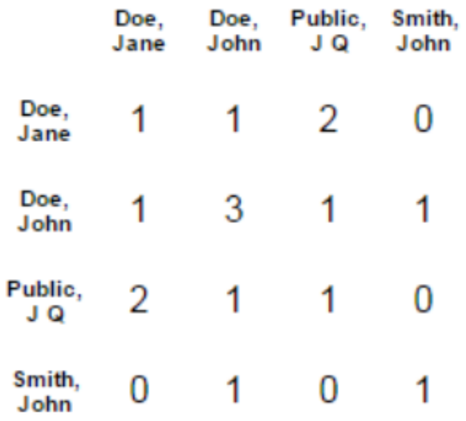
Technical Test – EMBL-EBI Position EBI_00816

Please complete the following three questions and email your response to applications@ebi.ac.uk with the reference number above in the subject line.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<MedlineCitationSet>
    <Article>
        <ArticleTitle>Title 1</ArticleTitle>
        <AuthorList>
            <Author>
                <LastName>Public</LastName>
                <ForeName>J Q</ForeName>
                <Initials>JQ</Initials>
            </Author>
            <Author>
                <LastName>Doe</LastName>
                <ForeName>John</ForeName>
                <Initials>J</Initials>
            </Author>
        </AuthorList>
    </Article>
    <Article>
        <ArticleTitle>Title 2</ArticleTitle>
        <AuthorList>
            <Author>
                <LastName>Doe</LastName>
                <ForeName>John</ForeName>
                <Initials>J</Initials>
            </Author>
            <Author>
                <LastName>Doe</LastName>
                <ForeName>Jane</ForeName>
                <Initials>J</Initials>
            </Author>
        </AuthorList>
</Article>
<Article>
        <ArticleTitle>Title 3</ArticleTitle>
        <AuthorList>
            <Author>
                <LastName>Doe</LastName>
                <ForeName>Jane</ForeName>
                <Initials>J</Initials>
            </Author>
            <Author>
                <LastName>Public</LastName>
                <ForeName>J Q</ForeName>
                <Initials>JQ</Initials>
            </Author>
        </AuthorList>
    </Article>
    <Article>
        <ArticleTitle>Title 4</ArticleTitle>
        <AuthorList>
            <Author>
                <LastName>Smith</LastName>
                <ForeName>John</ForeName>
                <Initials>J</Initials>
            </Author>
            <Author>
                <LastName>Doe</LastName>
                <ForeName>John</ForeName>
                <Initials>J</Initials>
            </Author>
        </AuthorList>
    </Article>
</MedlineCitationSet>
```



	Doe, Jane	Doe, John	Public, J Q	Smith, John
Doe, Jane	1	1	2	0
Doe, John	1	3	1	1
Public, J Q	2	1	1	0
Smith, John	0	1	0	1

a). The XML above lists four articles by four different authors. Your task is to generate a matrix similar to the one shown above. Each cell i, j should list the number of articles (say N) co-authored by the author in row i and the author is column j. You may output your results to the command line as a textual representation.

b). How would you design a test suite for the above application? You are not required to write any test cases but to describe your processes.

c). The above XML format is very similar to the Medline dataset, which contains over 22 million records. What problems do you foresee in scaling the above solution to these numbers of records.