

# Comparative Programming Languages

## UCT CSC304 2003 Practical Assignment

An inverted file is a mechanism used in search engines to support efficient query evaluation using pre-computed data structures. Inverted files are best illustrated with an example.

Suppose we have three documents (with names and content) as follows:

Document ID	List of words
doc1	apples oranges pears
doc2	oranges bananas
doc3	apples bananas oranges

The corresponding inverted files will then be:

Word	Inverted file contents
apples	doc1 doc3
bananas	doc2 doc3
oranges	doc1 doc2 doc3
pears	doc1

Suppose these inverted files are stored by a search engine. If a user supplies a query for “apples”, the contents of the inverted file reflect the list of search results – i.e. doc1 and doc3. If a user queries “apples bananas”, a typical search engine will compute an intersection of inverted files, resulting in the single document “doc3”, which is the only one that contains the supplied terms.

This assignment requires that you implement programs to compute inverted files from a set of given documents in each of the four programming paradigms:

1. Imperative – in ANSI C
2. Object-oriented – in Java
3. Declarative – in Prolog
4. Functional – in Clean

In each case, you need to write a program/class containing a “compute\_inverted” procedure/rule/function/method. Generate output as appropriate (screen snapshots, file dumps, etc.) to prove that your program works – the person marking your assignment will not necessarily read or execute your code so, if there is no output, it will be assumed that the program does not work!

Use the example above as test data – you do not need to use files but should instead use lists of words. For example, in Prolog you can define the first document as:

```
document(doc1, [apples, oranges, pears]).
```

Similarly, in Java this can be defined as:

```
doc1 = new document ("doc1", "apples oranges pears");
```

You may hard-code this test data in your programs provided that the subprograms to compute the inverted files are generic.