

## Named Entity Disambiguation:

Ciddhi Jain  
IIT Delhi

### Overview of problem:

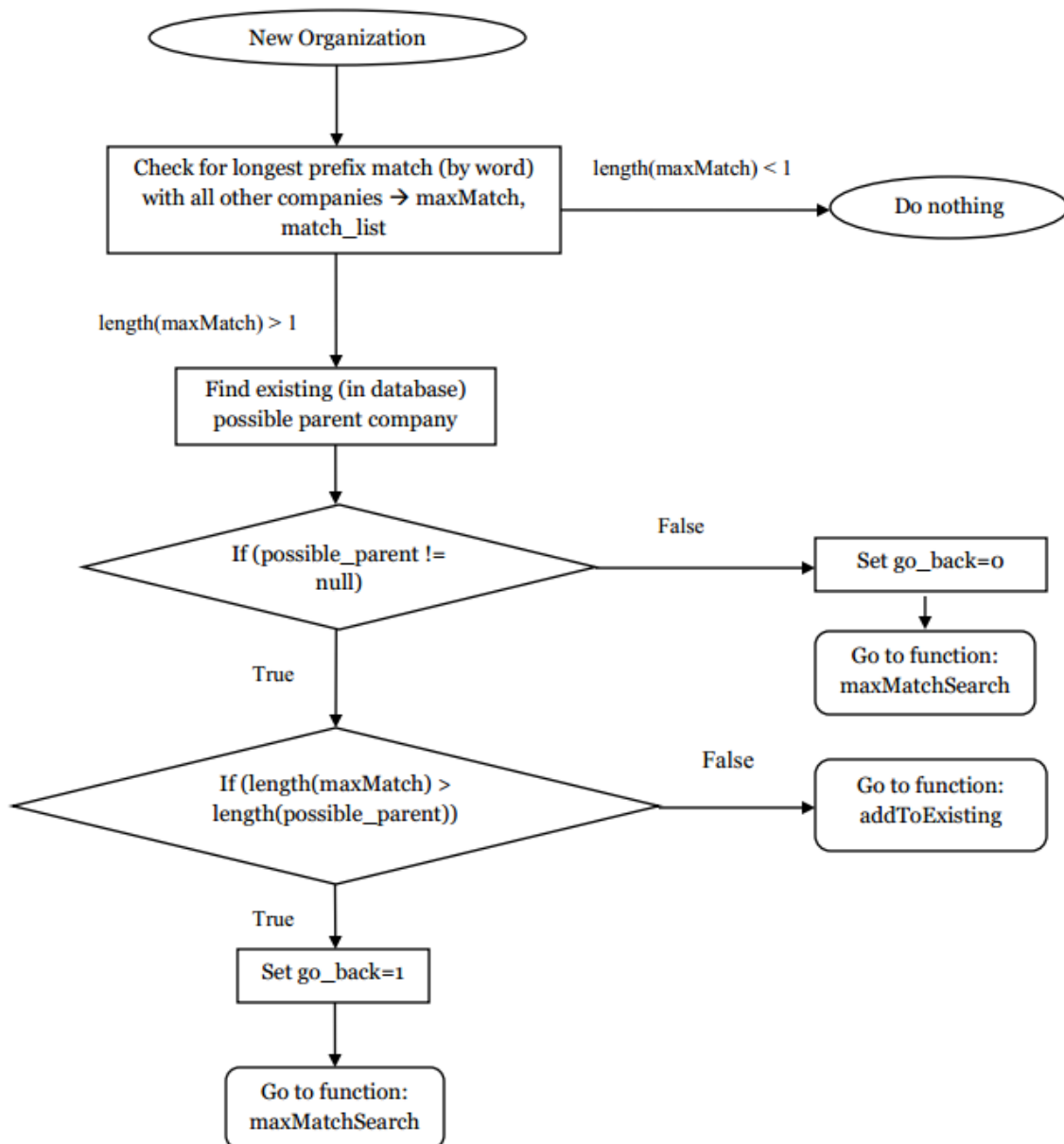
The following described code was developed for Info Assembly, a Financial **Big Data company** based in New Delhi and USA (ARK Challenge winner 2013). For a given list of organizations and companies (in orders of tens of thousands), the problem lies in associating subsidiaries to their parents. For example - Bloomberg News, Bloomberg Equity, Bloomberg India and all of its subsidiaries should be mapped to Bloomberg L.P. Also, Central Bank of Nigeria and Central Bank of Iran should not be considered as same.

### Proposed Solution:

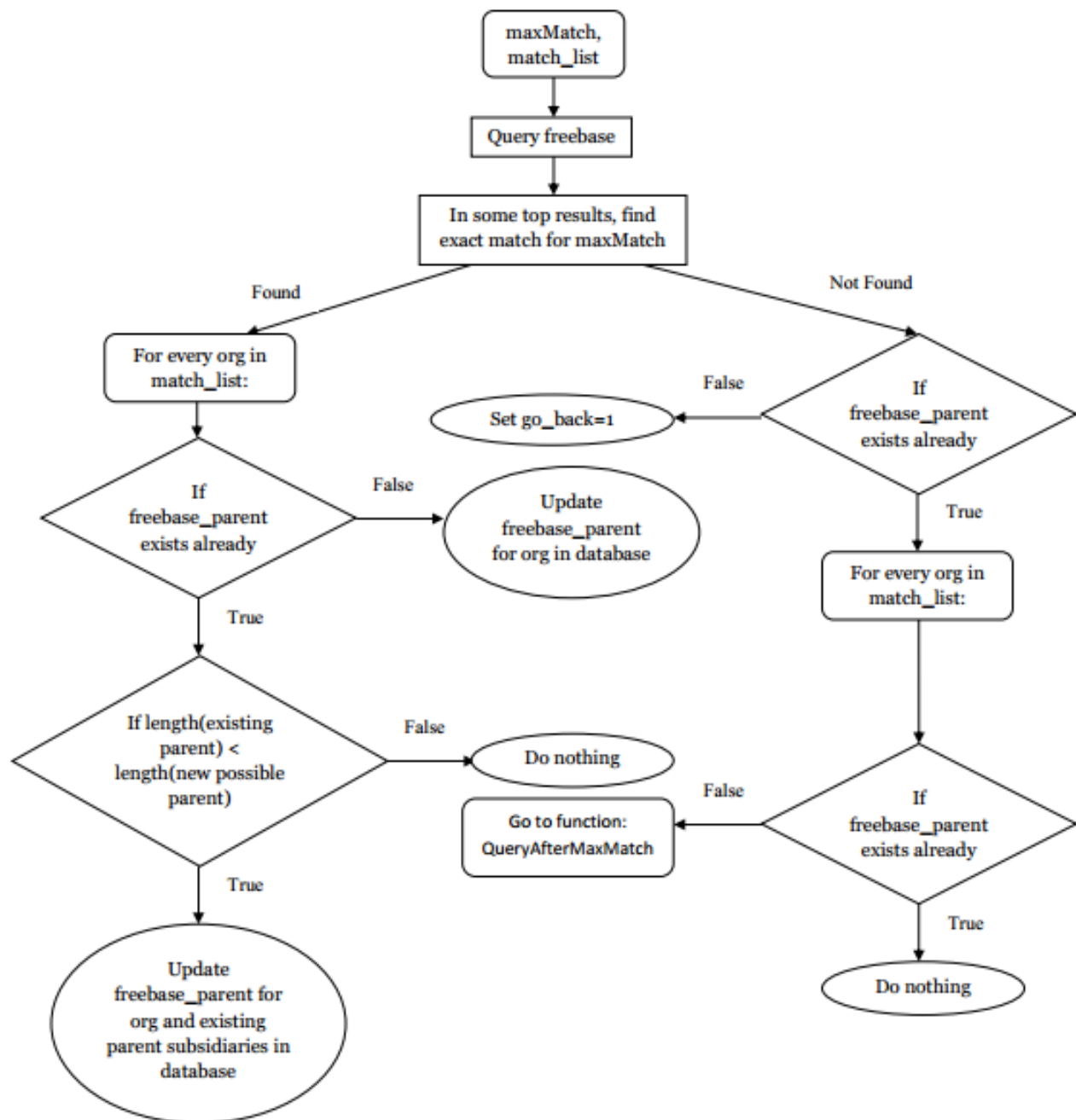
The broad solution lies in searching for **maximum matching prefix, not by letters but by words**. The common set of words is then queried on freebase and if an exact match is found, then all organizations sharing that match can be called as subsidiary of that parent found on freebase. For example - Goldman Sachs Global Holdings L.L.C., Goldman Sachs Asset Management, L.P., Goldman Sachs Hedge Fund Strategies LLC and all other subsidiaries of Goldman Sachs will be mapped to it and a parent-subsidary relationship will be maintained in the database. Also, any hypothetical company/organization named Goldman XYZ will not be marked as subsidiary. The following flowcharts describe exactly how the code performs Named-Entity Disambiguation on the large set and ensures that false mappings are minimized.

### Flowcharts:

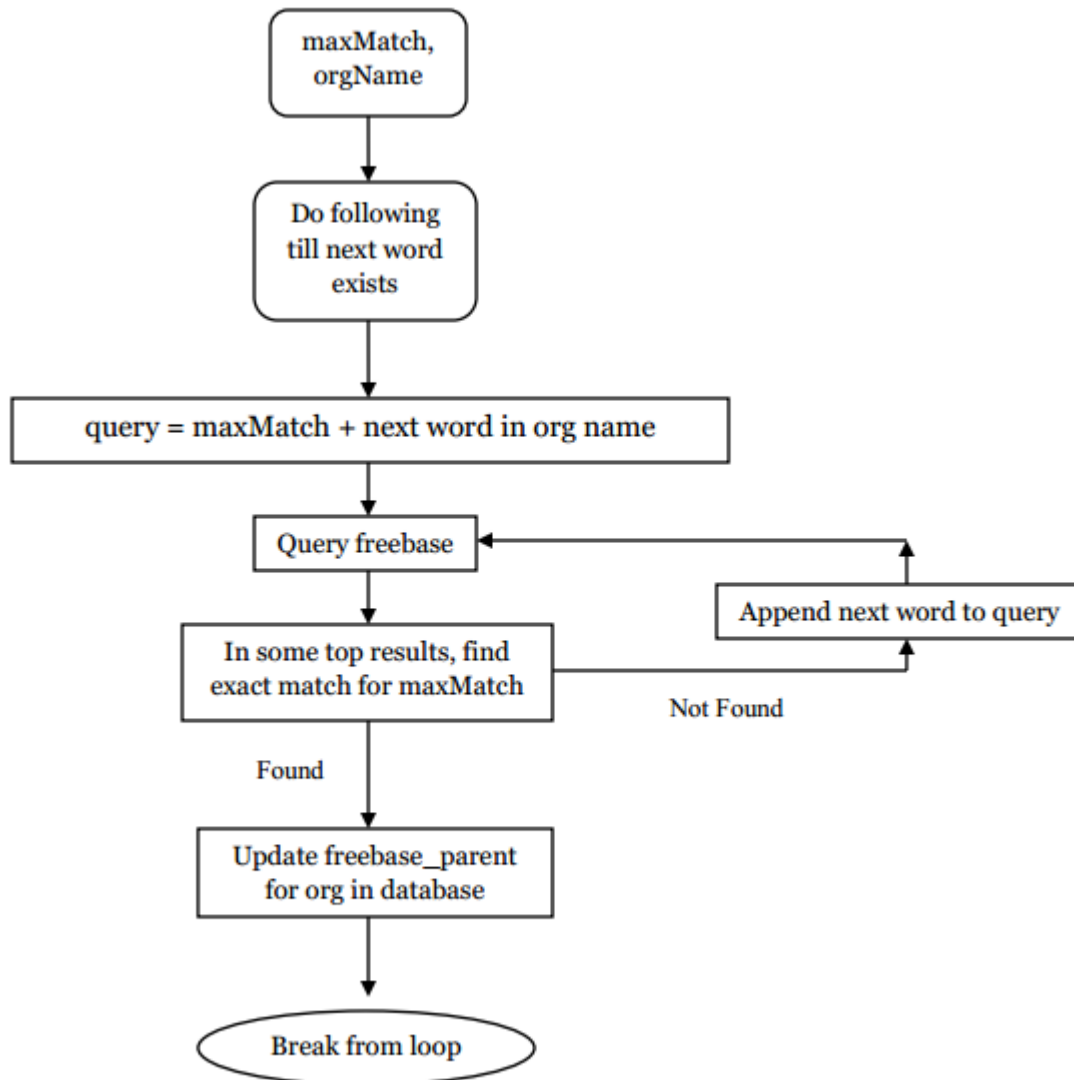
- **OrgsUtil.processOrgs(DBObject)** → This is the main function called to perform NED on a new organization entered in the Database. It in turn calls other functions whose flowcharts are also shown below.



- OrgsUtil.maxMatchSearch(maxMatch, match\_list)



- OrgsUtil.QueryAfterMatchSearch(maxMatch, orgName)



- OrgsUtil.AddToExisting(existingParent, orgName, orgsList)

