

**Information Retrieval Project phase 1**

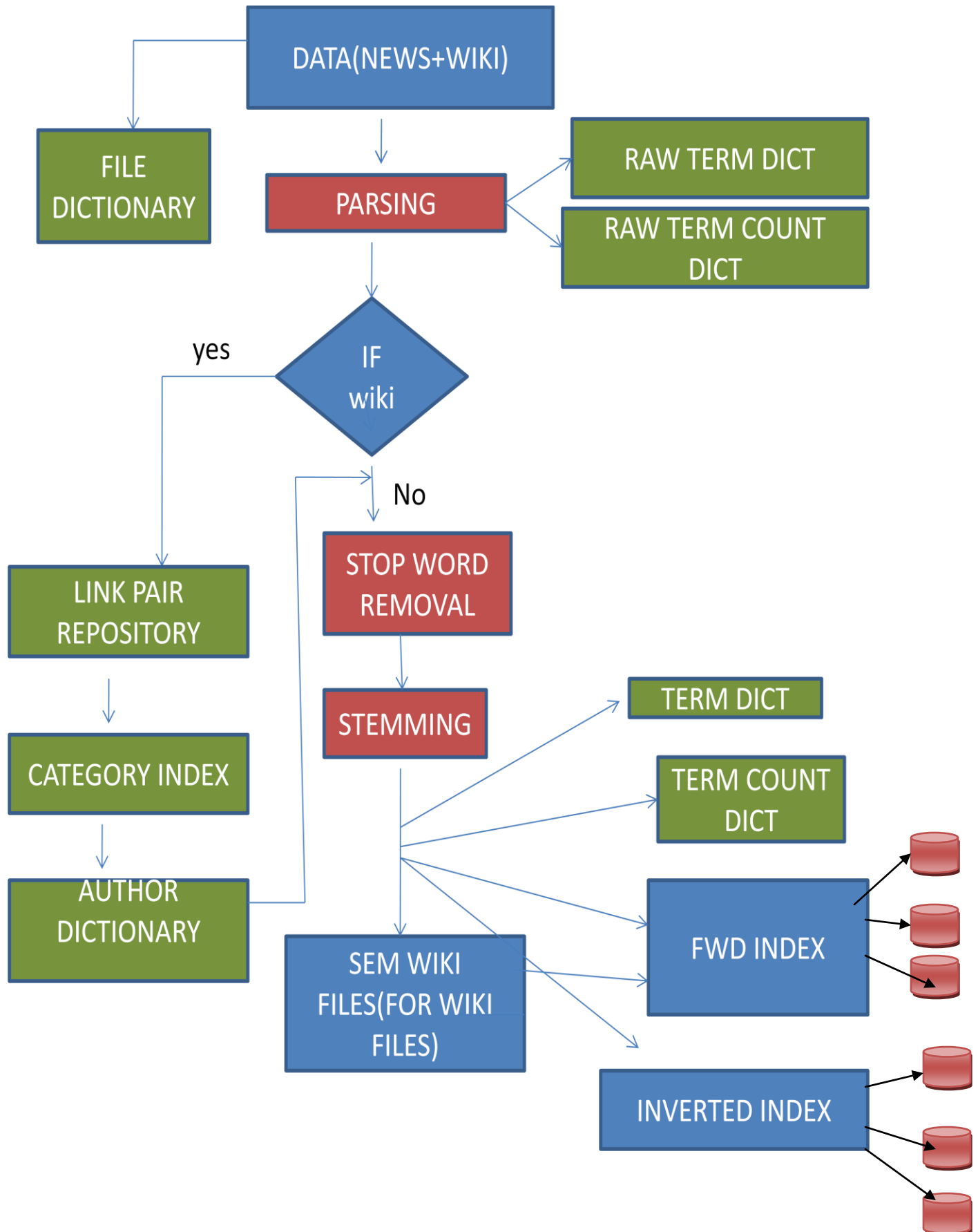
**TEAM –YAGOO**

**Dinesh Ravi**

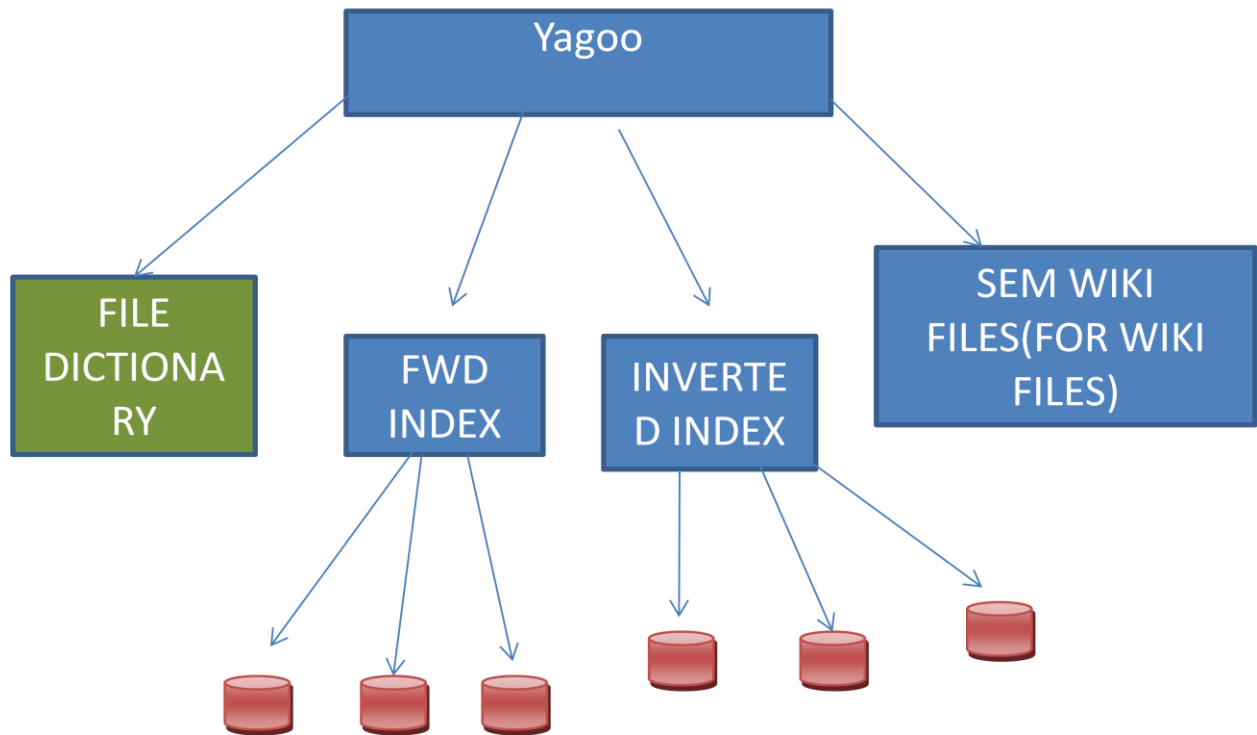
**Magizharasu Thirunavukkarasu**

**Manoj Mylapore Chandrasekaran**

# SYSTEM DESIGN DIAGRAM



## DIRECTORY STRUCTURE



### STATISTICS:

**File count : 149118**

**Raw word count : 165,229,178**

**Word count : 5,973,376**

**Sem wiki files created : 95642**

```
Applications Places System ? S
Mon Sep 26, 9:57 PM magi

magi@ubuntu: ~
File Edit View Terminal Help

dp->d_name = Адыгея, Республика.txt
dp->d_name = Вебер_К.И.txt
dp->d_name = Європейська партія України.txt
dp->d_name = Інтернет Партія України.txt
dp->d_name = Карачаево-Черкесия.txt
dp->d_name = Митрофанов.txt
dp->d_name = Сокольническая линия.txt
dp->d_name = Српска Радикална Странка Република Српска.txt
dp->d_name = Стандарт.txt
dp->d_name = Чрезвычайная комиссия.txt
dp->d_name = Т.txt
dp->d_name = 三菱ふそうトラック・バス株式会社.txt
dp->d_name = 中印关系.txt
dp->d_name = 倍賞美津子.txt
dp->d_name = 刺芹菇.txt
dp->d_name = 印度.txt
dp->d_name = 四禮使覽.txt
dp->d_name = 弘仁.txt
dp->d_name = 朱踐耳.txt
dp->d_name = 桑田佳祐.txt
dp->d_name = 浦島太郎.txt
dp->d_name = 没有共产党就没有新中国.txt
dp->d_name = 神宮球場.txt
dp->d_name = 謝選鋒.txt
dp->d_name = ルフィの過去！赤髪のシャンクス登場.txt
dp->d_name = 金荷言.txt
dp->d_name = 陳子玲.txt
dp->d_name = 鶴の木駅.txt
dp->d_name = 黃芪.txt
../../cse535/MAGI/index.fwd.8

File count149118

Raw word count=165229178
Words5973376

No of sem wiki files created: 95642
^[[B^[[B^[[D^[[D^[[D4.96237e+07metallica {~/new/Debug} >
```

## **CONTRIBUTION OF EACH MEMBER:**

**Dinesh Ravi : News files processing , Stemming , term, term count dictionaries**

**Magizharasu Thirunavukkarasu : Wiki parser , Sem wiki creation, Author Dictionary, Category index,Link pair repo.**

**Manoj Mylapore Chandrasekaran: Stopword, Raw text parsing , Index barrels .**

**Signature :**

**Dinesh Ravi**

**Magizharasu**

**Manoj**