FACILITATING WEB SERVICE DISCOVERY USING CLUSTERING

GUIDED BY : MRS SOWMYA KAMATH
PRESENTED BY
APOORVA CHANDRA(11IT09)
DILIP MALLYA(11IT19)
JNVK CHAITANYA(11IT32)
R.KAUSHIK(11IT65)

FOR MID EVALUATION OF WEB SERVICES(IT466)

PAPER IMPLEMENTED

- Web Services Discovery based on Semantic Similarity Clustering

- Authors : P.Ravinder Reddy, Damodaram.A

FEATURES OF THE SERVICES USED FOR CLUSTERING

```xml
<message name="doHello">
  <part name="yourName" type="tns:getName" />
</message>

<message name="doHelloResponse">
  <part name="return" type="tns:HelloResponse" />
</message>

<portType name="HelloPort">
  <operation name="doHello">
    <input message="tns:doHello" />
    <output message="tns:doHelloResponse" />
  </operation>
</portType>
```

PHPEveryDay.Com

```csharp
public class CalculatorWebService : System.Web.Services.WebService
{

    [WebMethod(MessageName="Add2Numbers")]
    public int Add(int firstNumber, int secondNumber)
    {
        return firstNumber + secondNumber;
    }


    [WebMethod]
    public int Add(int firstNumber, int secondNumber, int thirdNumber)
    {
        return firstNumber + secondNumber + thirdNumber;
    }
}
```
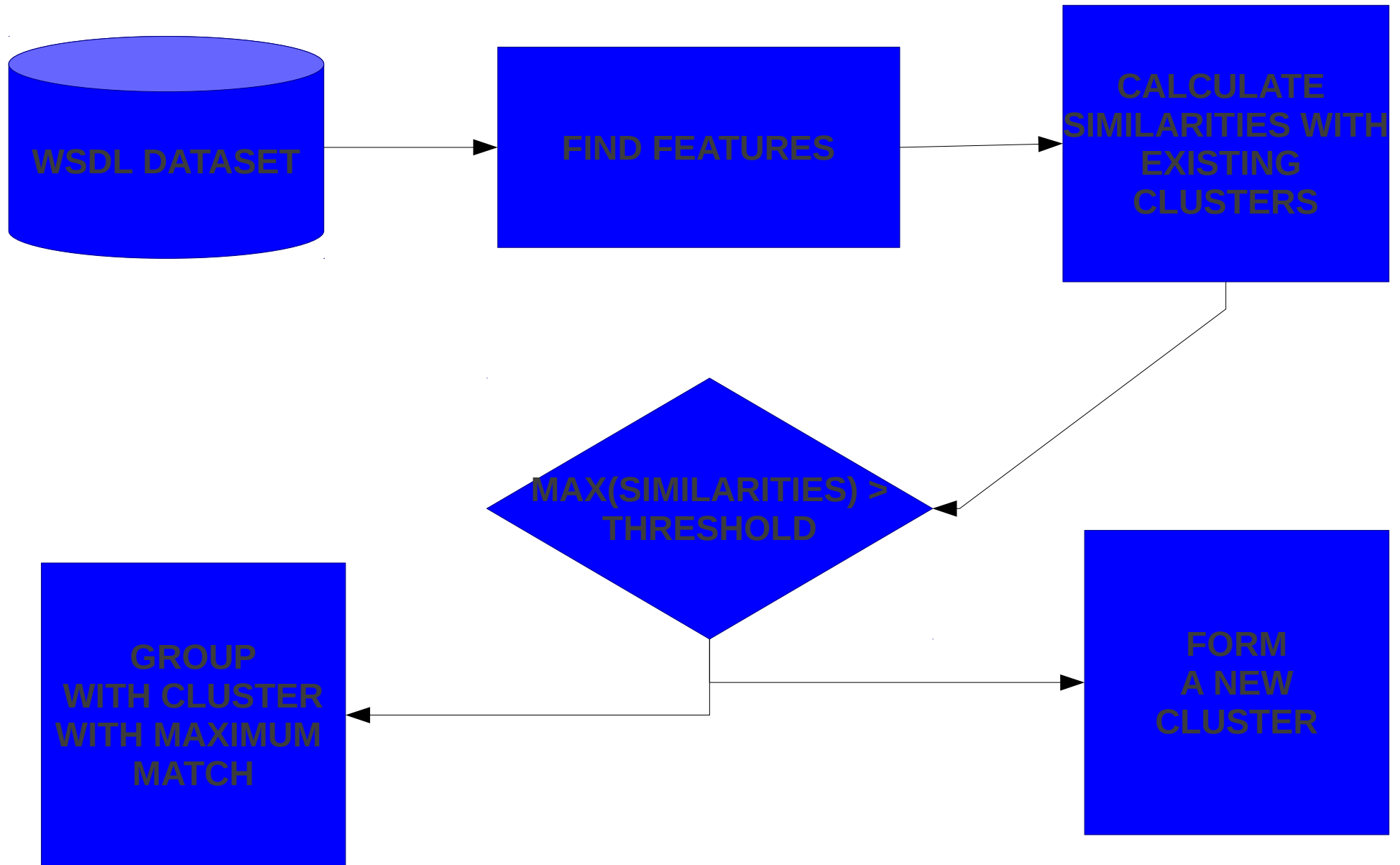
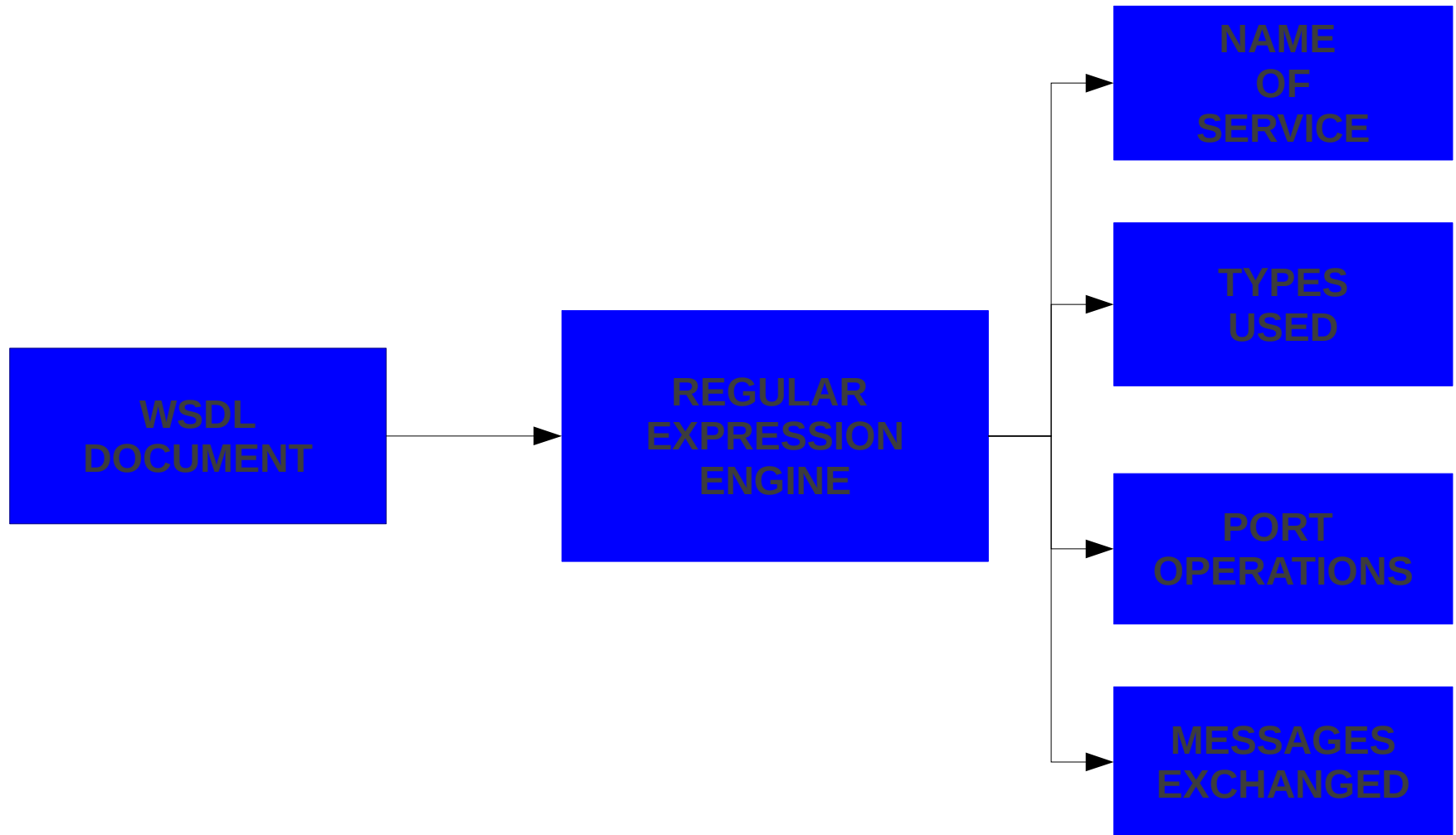# FEATURES OF THE SERVICES USED FOR CLUSTERING
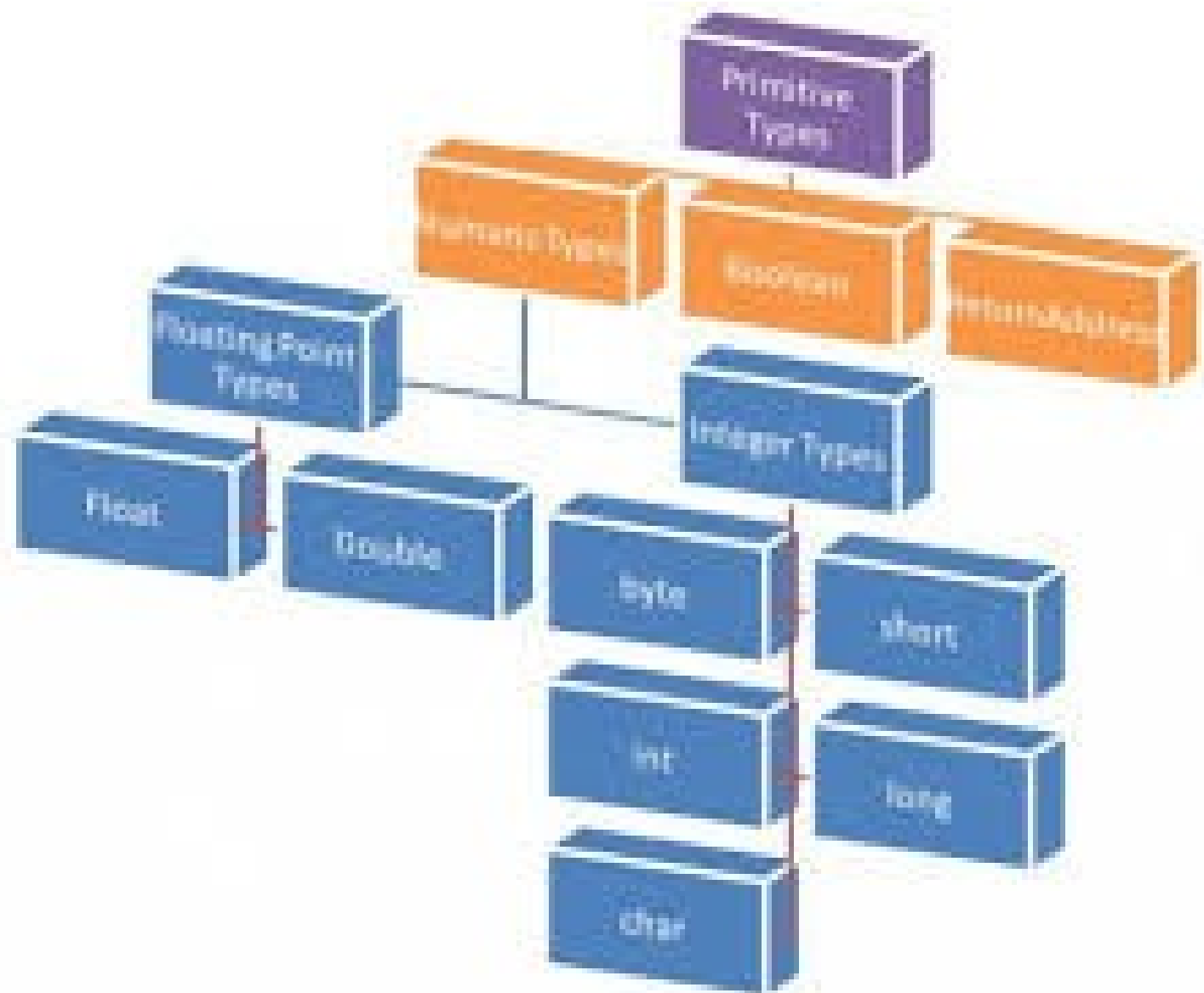
APPROACH USED

- Find similarity between the documents using the formula :

- Similarity = name_sim + mesg_sim + type_sim + port sim + tag_sim

- POINT TO BE NOTED : EQUAL WEIGHTS FOR ALL SIMILARITIES

WORKFLOW FOLLOWED

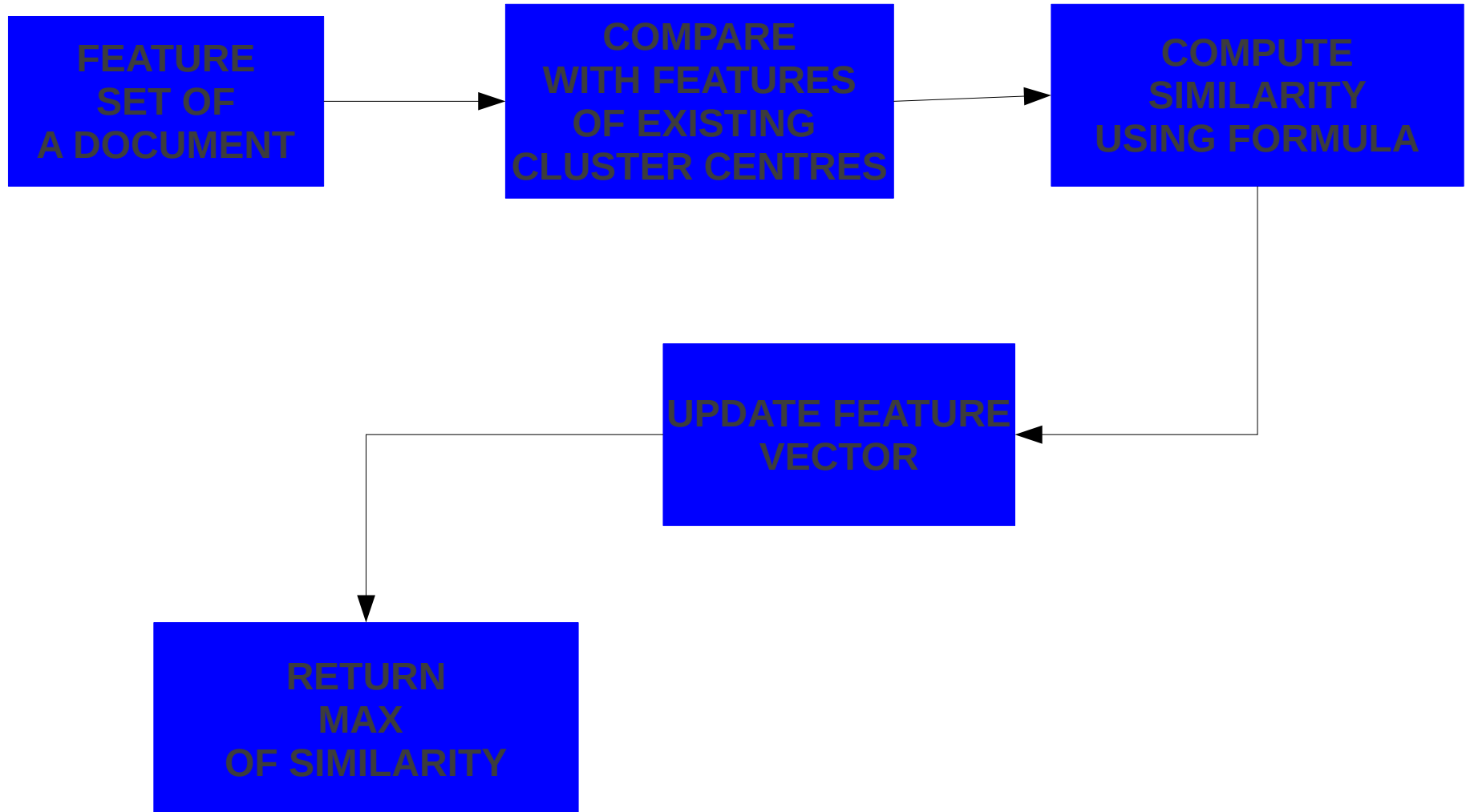WSDL DATASET → FIND FEATURES → CALCULATE SIMILARITIES WITH EXISTING CLUSTERS

MAX(SIMILARITIES) > THRESHOLD

GROUP WITH CLUSTER WITH MAXIMUM MATCH

FORM A NEW CLUSTER

FIND FEATURES

SIMILARITY WITH EXISTING CLUSTERS

```
FEATURE          →    COMPARE              →    COMPUTE
SET OF                WITH FEATURES             SIMILARITY
A DOCUMENT            OF EXISTING               USING FORMULA
                     CLUSTER CENTRES
```

```
                     UPDATE FEATURE
                     VECTOR
```

```
RETURN
MAX
OF SIMILARITY
```

NAME SIMILARITY FORMULA

- name_sim = lexical similarity(d1.name,d2.name)
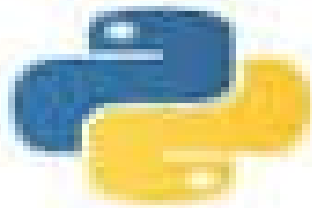
FINDING NAME SIMILARITIES

- Usage of nltk english dictionary
- Camel csae based separation
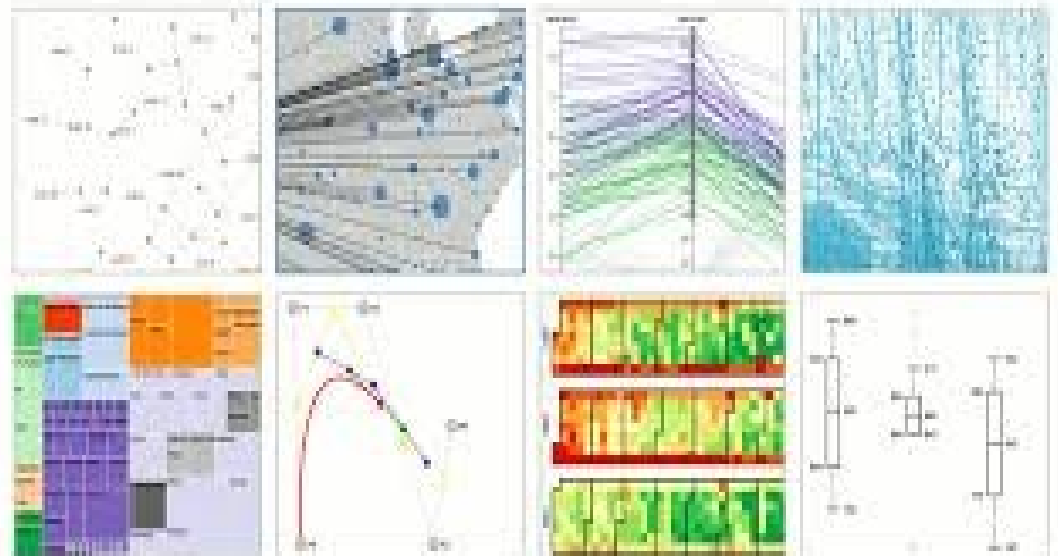
COMMON SIMILARITY FORMULA

- The common formula for finding similarity between the messages exchanged, port types of the services and types used by the services is :

- Sim = 2*match(l1,l2)/(len(l1)+len(l2))

TOOLS USED

OUR ADDITION TO THE FORMULA – UNEQUAL WEIGHTS

UNEQUAL WEIGHT VALUES

- Message names = 2.5(**most relevant**)
- Types used = 0.5(**least relevant**)
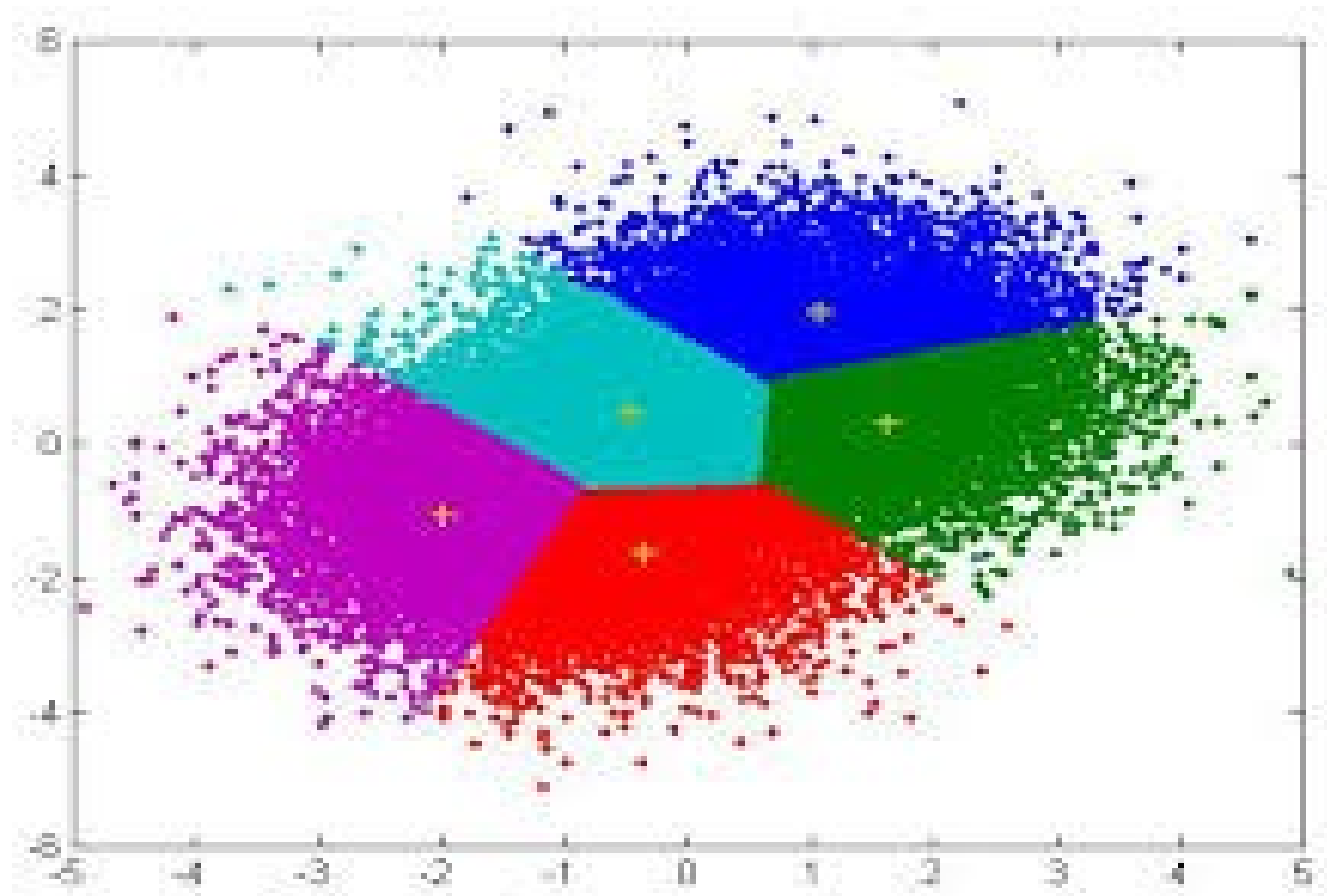- Message types = 1
- Port similarities = 1

FINAL FORMULA

- Similarity = name_sim + 2.5*mesg_name_sim + message_type_sim + 0.5*type_sim + port sim + tag_sim
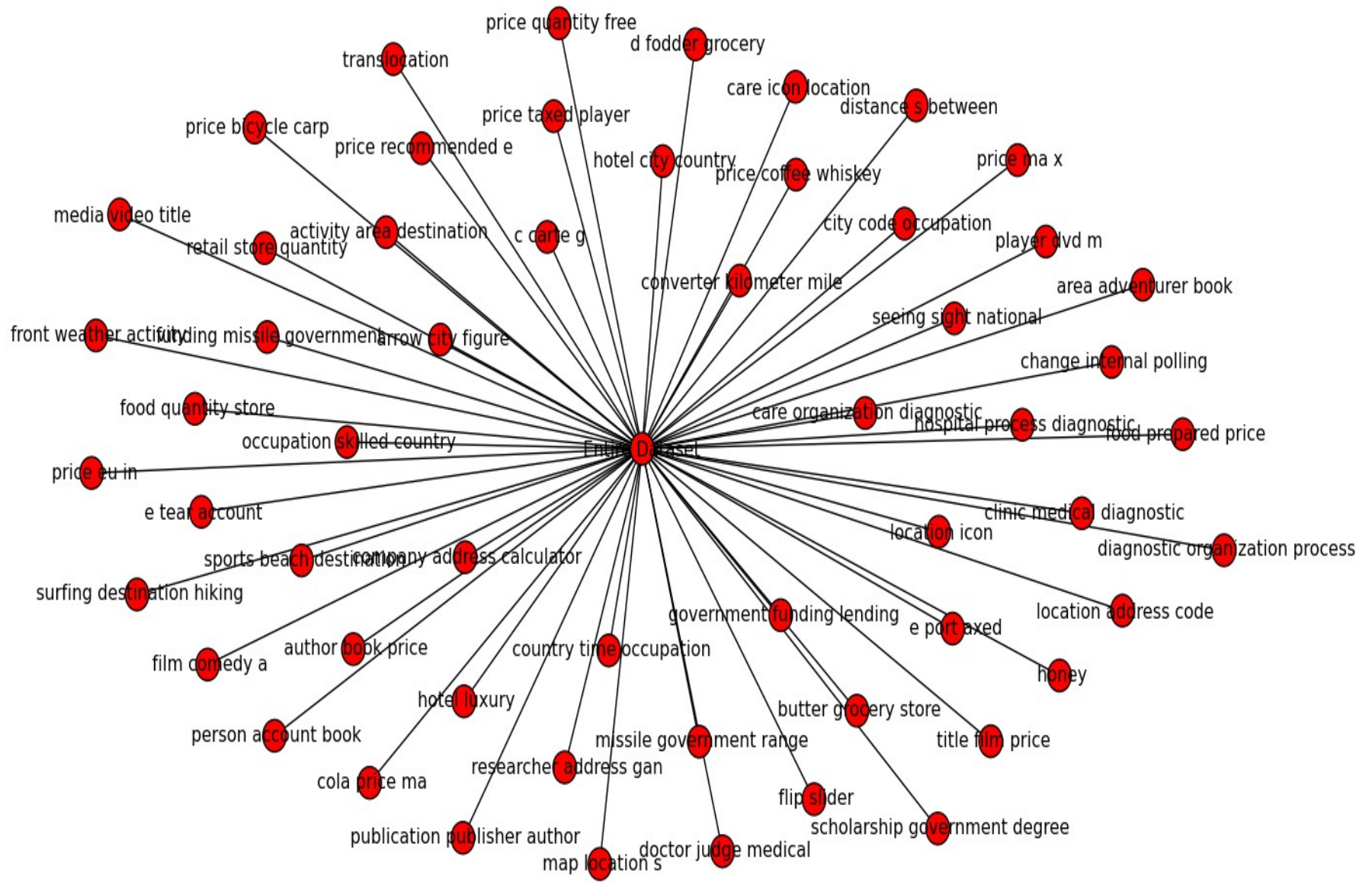
GROUPING OF DOCUMENTS

- Once a similarity score is obtained documnets will be clustered based on a threshold :

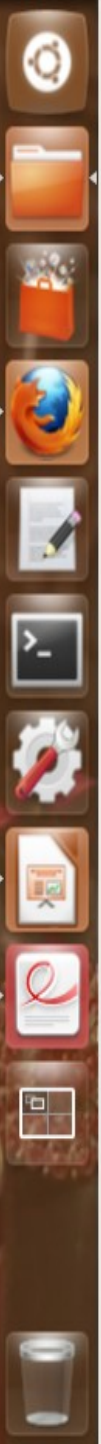- Assumed threshold = 0.9

FINAL RESULT AFTER THRESHOLD

RESULTS

- Of a dataset consisting of 1076 wsdl documents 56 clusters were obtained

- The name of the clusters has been given according to highest frequency of the words comprising the names of services in each cluster

Entire Dataset

price quantity free
d fodder grocery
translocation
care icon location
distance s between
price taxed player
price bicycle carp
price recommended e
hotel city country
price coffee whiskey
price ma x
media video title
activity area destination
c cane g
city code occupation
player dvd m
retail store quantity
converter kilometer mile
area adventurer book
front weather actifity
funding missile government
arrow city figure
seeing sight national
change internal polling
food quantity store
care organization diagnostic
hospital process diagnostic
food prepared price
occupation skilled country
price eu in
clinic medical diagnostic
location icon
diagnostic organization process
e tear account
sports beach destination
company address calculator
location address code
government funding lending
e port axed
surfing destination hiking
honey
author book price
country time occupation
film comedy a
hotel luxury
butter grocery store
person account book
missile government range
title film price
cola price ma
researcher address gan
flip slider
scholarship government degree
publication publisher author
doctor judge medical
map location s

✉ ▭ 1:32 ◇ ◀)) Fri Oct 10 7:17 PM ⚙

Bookmarks
🔒 x-nautilus-de...

⬩ 🏠 Home | 🖥 Desktop | web services project | **Clusters**

‹ › Q Search

Computer
🏠 Home
🖥 Desktop
📄 Documents
⬇ Downloads
🎵 Music
🖼 Pictures
🎬 Videos
💾 File System
🗑 Trash

Network
🖥 Browse Net...

perso count cityc durat	_hote FoodR Autho Expen	AutoR Coffe bicyc Title	4whee ZAD_s store SPD-G	prepa hikin surfi hikin	surfi FoodM meat_ retai	meat_ medic medic medic
cluster0.txt	cluster1.txt	cluster2.txt	cluster3.txt	cluster4.txt	cluster5.txt	cluster6.txt
medic BookA novel Book.	BookA Count getMa real-	Count title title title	_vide gover natio balli	gover weapo missi Missi	missi resea resea unive	Check Activ CityA CitvW
cluster7.txt	cluster8.txt	cluster9.txt	cluster10.txt	cluster11.txt	cluster12.txt	cluster14.txt
Check gover missi balli	flip_	agent print publi Perso	print Cycle shopp Recon	Shopp locat shopp Recom	fall_ natio degre gover	Journ hospi Hospi hospi
cluster16.txt	cluster17.txt	cluster18.txt	cluster19.txt	cluster20.txt	cluster23.txt	cluster25.txt
hospi retai Corpo retai	Retai title Title title	Title Carea Carea Carea	Carea locat	locat getLo check querv	findN surfi sport sport	sport Quali Maxpr Maxpr
cluster26.txt	cluster27.txt	cluster28.txt	cluster29.txt	cluster30.txt	cluster31.txt	cluster32.txt
untan retai merca groce	_food getDi usPos getDi	addre compa Count	Compa hotel Price taxfr	switc Groce	_fodd Activ Activ Activ	Activ Weath Weath CitvW
cluster33.txt	cluster34.txt	cluster35.txt	cluster36.txt	cluster40.txt	cluster41.txt	cluster42.txt
_weat Coffe Cdpla 4whee	_prop 4whee CarTe Fastc	_hone	Check googl stati qetMa	Selec scien monog Encvc	_icin Camer tea_t Butte	_cere mp3pl Dvdpl Dvdpl
cluster43.txt	cluster45.txt	cluster48.txt	cluster53.txt	cluster55.txt	cluster61.txt	cluster63.txt
_poll Inter	fill_ organ Unsuc Diagn	_weat Comed Comed	GetNa tea_r _coff Drugs	Ebook surfi Adven	_tran	Bookm Groce groce
cluster66.txt	cluster69.txt	cluster73.txt	cluster76.txt	cluster81.txt	cluster82.txt	cluster84.txt
Provi surfi sight siqht	locat	flip_ car_r CarTa Drouq	redLi count Count Count	mileT	_lux	_medi _USju
cluster87.txt	cluster88.txt	cluster94.txt	cluster97.txt	cluster102.txt	cluster106.txt	cluster109.txt

WORK REMAINING
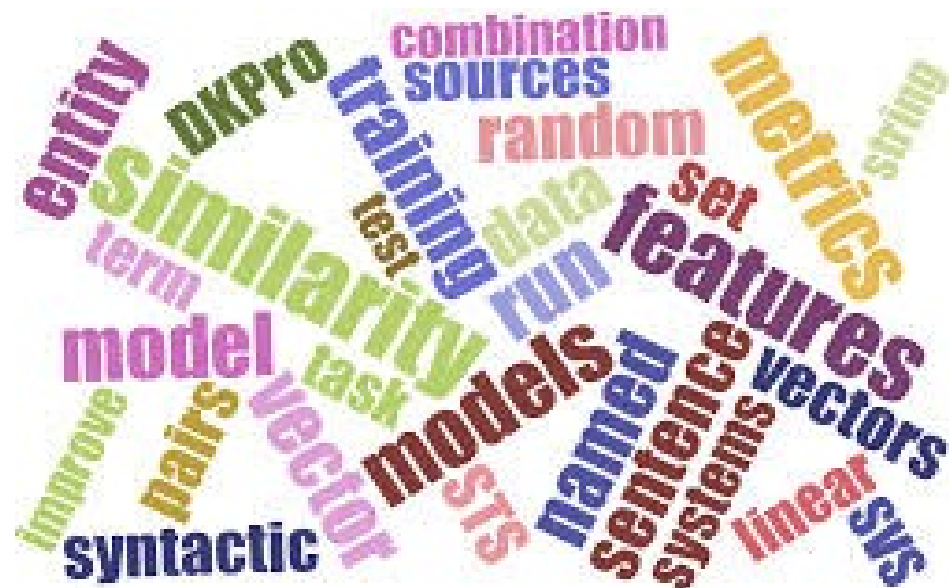
THE FINAL PRODUCT

Service-Finder

QUESTIONS