For compiling, the following command was used:

mpicc -Wall -g -fopenmp -o mpi\_word\_embeddings ./mpi\_word\_embeddings.c

The Wall and g flags are there to assist in debugging and the library OpenMP is used to parallelize loops.

For running the command was:

mpirun -n 11 --oversubscribe mpi\_word\_embeddings

The oversubscribe flag allowed me to run the program with few cores in my computer, the n with value 11 means 1 master and 10 slave nodes (P in the instructions).

The tags correspond to each kind of data that was sent with MPI messages, they were defined with #define because their values are constant, globally used and their addresses are not needed.

The commands are defined as const ints, as their addresses are needed, as the first parameter of MPI\_Send is a pointer.

A maximum word length of 1024 was considered enough, as no word from the file is bigger than that. The scanf is limited to 1023 input characters so that there’s space for the ending \0.

Doubles were used to store weights for their better accuracy over floats, especially since addition is used on them. However, in situations with more data, a float could be a more sensible option, as vector operations can perform faster on them and they need less space.

Any loop that iterated over all the slave processes had the #pragma omp for pragmas so that they would perform in parallel since they didn’t mutate any shared data.

The rows of a process were calculated with the formulas:

first row (inclusive) = (rank-1) × c + min (rank-1, r)

and

last row (exclusive) = rank × c + min (rank, r).

This allows us to split evenly the rows and to determine the size of the matrix needed for each one (and thus only allocating that size on the main process when reading the file) with *range = last row – first row.*

The EXIT command just sends it to the slave nodes and they stop their execution.

The QUERY command sends two messages from the root process, one for the command and another for the word. This allows separation of the command selection logic from the query code itself.

Then also two messages are received but just one if the process didn’t have the word. That way only the process with the word sends another message, reducing overhead.

If no slave node knows the word, the query ends there with an error message.

If VERSION\_TOP\_P is defined, then the code for getting the most relevant words is enabled.

In case a word is found, the command CALCULATE\_SIMILARITY is used so that each slave process finds the most similar word. This is called just once if VERSION\_TOP\_P is not defined, if it is, then a merge algorithm is used, where the most similar word from the processes is selected and the corresponding process sends the next most similar one (using another CALCULATE\_SIMILARITY command). For this, each slave node stores a bitmap (for simplicity, a byte was used to store each boolean value) of the already sent words and omits those in the comparisons. The bitmap is reset on a new QUERY command but not on another CALCULATE\_SIMILARITY command.

## **Found results:**

Note: Some of the words for the example query (telefon) aren’t found in the word embeddings file.

### **boğaziçi:**

akyaka with score 0.551848 from process 1

altunizade with score 0.559370 from process 2

ayazağa with score 0.623934 from process 3

boğaziçi with score 1.000000 from process 4

galata with score 0.592010 from process 5

istanbul with score 0.630798 from process 6

marmara with score 0.639540 from process 7

rumelihisarı with score 0.644971 from process 8

tarabya with score 0.598414 from process 9

üsküdar with score 0.589914 from process 10

**Most relevant words:**

boğaziçi with score 1.000000

rumelihisarı with score 0.644971

marmara with score 0.639540

odtü with score 0.635022

istanbul with score 0.630798

ayazağa with score 0.623934

bilkent with score 0.612875

beyazıt with score 0.598733

kağıthane with score 0.598623

tarabya with score 0.598414

### **üniversite:**

akademisyen with score 0.672107 from process 1

anaokul with score 0.573987 from process 2

açıköğretim with score 0.647897 from process 3

burs with score 0.716351 from process 4

fakülte with score 0.838503 from process 5

kolej with score 0.732595 from process 6

mezun with score 0.708246 from process 7

rektör with score 0.784243 from process 8

yök with score 0.666313 from process 9

üniversite with score 1.000000 from process 10

**Most relevant words:**

akademisyen with score 0.672107 from process 1

anaokul with score 0.573987 from process 2

açıköğretim with score 0.647897 from process 3

burs with score 0.716351 from process 4

fakülte with score 0.838503 from process 5

kolej with score 0.732595 from process 6

mezun with score 0.708246 from process 7

rektör with score 0.784243 from process 8

yök with score 0.666313 from process 9

üniversite with score 1.000000 from process 10

### **bilgisayar:**

adresle with score 0.603858 from process 1

anahtarla with score 0.561217 from process 2

bilgisayar with score 1.000000 from process 3

dizüstü with score 0.761838 from process 4

harddisk with score 0.697283 from process 5

internet with score 0.660916 from process 6

masaüst with score 0.697266 from process 7

pc with score 0.640958 from process 8

yazılım with score 0.740554 from process 9

çip with score 0.704136 from process 10

**Most relevant words:**

bilgisayar with score 1.000000

dizüstü with score 0.761838

yazılım with score 0.740554

çip with score 0.704136

harddisk with score 0.697283

masaüst with score 0.697266

masaüstü with score 0.691609

bellek with score 0.673514

modem with score 0.669376

internet with score 0.660916

### **mühendis:**

akademisyen with score 0.561000 from process 1

analitik with score 0.463634 from process 2

başmühendis with score 0.753364 from process 3

biyoteknoloji with score 0.560986 from process 4

imalât with score 0.558885 from process 5

jeofizik with score 0.608447 from process 6

mühendis with score 1.000000 from process 7

profesör with score 0.570566 from process 8

teknisyen with score 0.711158 from process 9

çalış with score 0.538000 from process 10

**Most relevant words:**

mühendis with score 1.000000

başmühendis with score 0.753364

teknisyen with score 0.711158

mühendishane with score 0.675721

bilim with score 0.614011

jeofizik with score 0.608447

uzman with score 0.605567

biyolog with score 0.596059

profesör with score 0.570566

müh with score 0.562473

### **bölüm:**

adanmak with score 0.419602 from process 1

altbölüm with score 0.578744 from process 2

başkarakter with score 0.550873 from process 3

bölüm with score 1.000000 from process 4

freakazoid with score 0.536178 from process 5

işbölüm with score 0.597805 from process 6

kısım with score 0.556643 from process 7

sezon with score 0.649103 from process 8

yasemin with score 0.503049 from process 9

yıl with score 0.468516 from process 10

**Most relevant words:**

bölüm with score 1.000000

bölümle with score 0.836784

bölümlemek with score 0.836784

sezon with score 0.649103

dizi with score 0.630106

işbölüm with score 0.597805

işbölümü with score 0.597805

jenerik with score 0.579716

altbölüm with score 0.578744

jeneriğ with score 0.570887

### **algoritma:**

aksiyom with score 0.568324 from process 1

algoritma with score 1.000000 from process 2

aritmetik with score 0.606194 from process 3

değişken with score 0.652004 from process 4

genelle with score 0.614804 from process 5

karmaşık with score 0.659805 from process 6

logaritma with score 0.697480 from process 7

polinom with score 0.661473 from process 8

teorem with score 0.632696 from process 9

özyinele with score 0.704810 from process 10

**Most relevant words:**

algoritma with score 1.000000

özyinele with score 0.704810

özyinelemek with score 0.704810

logaritma with score 0.697480

şifrele with score 0.663583

şifrelemek with score 0.663583

polinom with score 0.661473

karmaşık with score 0.659805

optimizasyon with score 0.654680

değişken with score 0.652004

### **analiz:**

aksiyom with score 0.616612 from process 1

analiz with score 1.000000 from process 2

araştır with score 0.670660 from process 3

doğru with score 0.657478 from process 4

genelle with score 0.653200 from process 5

ince with score 0.668979 from process 6

metot with score 0.676078 from process 7

sına with score 0.663874 from process 8

teori with score 0.655900 from process 9

ölçümle with score 0.673152 from process 10

**Most relevant words:**

analiz with score 1.000000

analizör with score 0.737437

metot with score 0.676078

ölçümle with score 0.673152

ölçümlemek with score 0.673152

araştır with score 0.670660

araştırmak with score 0.670660

ince with score 0.668979

sına with score 0.663874

sınamak with score 0.663874

### **ders:**

akademi with score 0.573470 from process 1

anaokul with score 0.661514 from process 2

açıköğretim with score 0.678602 from process 3

ders with score 1.000000 from process 4

eğitim with score 0.761102 from process 5

imtihan with score 0.626399 from process 6

okul with score 0.791339 from process 7

ortaokul with score 0.712584 from process 8

veli with score 0.644447 from process 9

öğretmen with score 0.778179 from process 10

**Most relevant words:**

ders with score 1.000000

okul with score 0.791339

derslik with score 0.780701

öğretmen with score 0.778179

dersliğ with score 0.766389

eğitim with score 0.761102

öğret with score 0.749407

öğretmek with score 0.749407

ortaokul with score 0.712584

ilkokul with score 0.708791

### **proje:**

aiesec with score 0.514102 from process 1

amaç with score 0.572893 from process 2

ar with score 0.615446 from process 3

destek with score 0.561457 from process 4

girişim with score 0.633173 from process 5

inşaat with score 0.606869 from process 6

konut with score 0.637415 from process 7

proje with score 1.000000 from process 8

yatırım with score 0.624813 from process 9

özel with score 0.585725 from process 10

**Most relevant words:**

proje with score 1.000000

konut with score 0.637415

girişim with score 0.633173

planlamak with score 0.629297

yatırım with score 0.624813

ar with score 0.615446

arsa with score 0.607970

inşaat with score 0.606869

geliştir with score 0.604534

geliştirmek with score 0.604534

### **ödev**

ahlak with score 0.504182 from process 1

alışkan with score 0.524277 from process 2

anla with score 0.527655 from process 3

ders with score 0.623479 from process 4

eğitim with score 0.550968 from process 5

iş with score 0.548328 from process 6

on with score 0.526729 from process 7

sorum with score 0.551424 from process 8

veli with score 0.576351 from process 9

ödev with score 1.000000 from process 10

**Most relevant words:**

ödev with score 1.000000

ödevle with score 0.947491

ödevlemek with score 0.947491

ders with score 0.623479

veli with score 0.576351

özür with score 0.554547

sorum with score 0.551424

öğret with score 0.551083

öğretmek with score 0.551083

eğitim with score 0.550968