

Assignment 4

HITS Algorithm

File Format

```
Ur11 Ur12 Ur13
Ur14 Ur12
Ur13 Ur11 Ur15 Ur16
Ur12 Ur11 Ur14 Ur16 Ur17
Ur15
.
.
.
.
Ur1N Ur11
```

Example

```
http://www.it.uu.se/index.html http://www.uu.se  
http://www.kth.se/ www.uu.se http://www.slu.se  
http://www.uu.se http://www.it.uu.se www.dn.se  
http://user.it.uu.se/~pergu http://www.uu.se  
www.slu.se http://www.uu.se http://www.liu.se  
http://www.it.uu.se/ http://www.slu.se  
http://www.liu.se
```

Example

Same URL different names

```
http://www.it.uu.se/index.html http://www.uu.se  
http://www.kth.se/ www.uu.se http://www.slu.se  
http://www.uu.se http://www.it.uu.se www.dn.se  
http://user.it.uu.se/~pergu http://www.uu.se  
www.slu.se http://www.uu.se http://www.liu.se  
http://www.it.uu.se/ http://www.slu.se  
http://www.liu.se
```

Example

Same URL different names

```
http://www.it.uu.se/index.html http://www.uu.se  
http://www.kth.se/ www.uu.se http://www.slu.se  
http://www.uu.se http://www.it.uu.se www.dn.se  
http://user.it.uu.se/~pergu http://www.uu.se  
www.slu.se http://www.uu.se http://www.liu.se  
http://www.it.uu.se/ http://www.slu.se  
http://www.liu.se
```

Example

Only linked to

```
http://www.it.uu.se/index.html http://www.uu.se  
http://www.kth.se/ www.uu.se http://www.slu.se  
http://www.uu.se http://www.it.uu.se www.dn.se  
http://user.it.uu.se/~pergu http://www.uu.se  
www.slu.se http://www.uu.se http://www.liu.se  
http://www.it.uu.se/ http://www.slu.se  
http://www.liu.se
```

HITS Algorithm

Idea:

Each webpage has an **authority** value and a **hubbusiness** value with respect to a query

Authority:

How “trustworthy” this page is with respect to the query

Hubbusiness:

How good a site is at linking to interesting sites for the query

HITS Algorithm

How do we calculate authority and hubbiness values?

Authority:

The sum of the hubbiness of the sites linking to this site

Hubbiness:

The sum of the authority of the site this site is linking to