**Clusters Search Engine Algorithm**

C. D. Jones

1. Divides searches into subject matter. (Like Google has “Images”, “News”, “Maps”
2. To do this the data mining algorithm has to be able to classify pages and sift through subject matter.
3. The user is not “searching” but asking the search engine questions to which the search engine answers via the pages it returns.

Properties of Cluster Algoirthm:

Common tags between two pages cause them to move together or “cluster”

Distance is a function of similar content.

Clusters are formed when individual pages have common terms. Entire clusters can be attracted to other clusters.

Large topics consist of major topics, while smaller clusters are subtopics.

Cluster space is defined using antonyms, where for example “hot” is on the opposite end of “Cold”. Antonym tags cause pages and clusters to repel.

**Example Clusters**

1.

<home>

Hi My Name is Ted and I’m building a search engine.

</home>

2.

<body>

Where is Joel? He’s late for his dentist appointment.

</body>

3.

<body>

I’m running a search on Joel in the search engine

</body>

Since the majority of movement is towards the origin, sending newly mined data to a pre-defined location representing the outskirts of the cluster space seems to make the most sense. The value for the outskirts boundary should be calculated empirically.

“It” Cluster

“Change” Cluster

Health Cluster

This data space can be useful for a different, more language based search engine.

Data Filter

Send to Cluster Space

Mine Data

Cluster Space

Search Engine Front End