Anastasia’s Präsentation ging um eine Websuche mit der Hilfe von zwei Algorithmen - PageRank und HITS. Zuerst hat Anastasia über den Inhalt gesprochen, nämlich die Definitionen der Algorithmen, ihre Vorteile und Nachteile, Berechnungsbeispiele und ihren Vergleich und Literatur.
Web ist ein Hypertext mit vielen Hilfsinformationen, welcher enthält Linkstruktur & Linktext. Klassische Suchmaschinen basieren auf einer textbasierten Analyse von Seiten und moderne Suchalgorithmen basieren sich auf das Matchen der Suchbegriffe mit einem großen Satz indizierter Webseiten. Das Matchen lässt sich durch eine Term-Dokumenten-Datei realisieren. Sie sind wie die Tabellen mit Tokens und dazugehörige Dokumente dargestellt.
Während der Suche bekommt man meistens extrem viele Treffer,die stark variieren. Somit sind Algorithmen nötig, welche die Relevanz von Webseiten für bestimmte Suchanfrage bewerten. Dafür braucht man die Algorithmen wie PageRank und HITS, die auf der Idee basieren, dass man Infos über die Relevanz durch die Betrachtung der Linkstruktur erhält.
PageRank wurde 1996 von Larry Page und Sergey Brin an der Stanford University entwickelt und damals für Suchmaschine Google verwendet. Es basiert auf der Idee dass die Anzahl der eingehenden Links als Maßstab der Relevanz ist und nicht alle Links sind gleichwertig, das heißt: eine Seite ist umso wichtiger, je mehr und wichtigere Seiten auf sie verlinken. Die Formel zu dem Algorithm besagt daß die Relevanz einer Seite durch die Anzahl und Relevanz der auf sie linkenden Seiten bestimmt wird und der Graph ist ist eine quadratische Matrix, deren Zeilen und Spalten den Internetseiten entsprechen. Aber es gibt ein Problem - Rang-Senke,es sind zwei Seiten, die auf sich gegenseitig und sonst auf keine weitere Seite zeigen und noch eine weitere Seite, die auf eine der beiden zeigt. Dazu gibt es eine Lösung - Rang-Quelle,daß ist ein Faktor für alle Seiten, der jeder Seite einen Wert als Rang-Quelle zuweist und ist auch als Vektor interpretierbar.
Anastasia hat auch ein Random Surfer Modell erwähnt. Hier PageRank erscheint intuitiv, wenn man das Random Surfer Modell zugrunde legt und startet bei einer Internetseite, dann klickt er wahllos auf irgendwelche Links und langweilt sich irgendwann und sucht eine andere Seite auf. Rang einer Internetseite entspricht hier der Wahrscheinlichkeit, dass sich der Random Surfer zu einem beliebigen Zeitpunkt gerade auf dieser Internetseite befindet.
Generell der größte Nachteil des PageRank ist dass die Berechnung extrem zeitaufwendig ist, allerdings nicht zur Zeit der Suchanfrage, sondern im Voraus, aber dafür arbeitet er völlig unabhängig von einer Suchanfrage.
HITS wurde 1997 von Jon Kleinberg entwickelt. Die Grundidee liegt daran daß die Anzahl der eingehenden Links als Maßstab der Relevanz ist und man betrachtet nur einen kleinen Teilgraphen des Internets und schreibt jedem Dokument darin zwei Werte zu, einen als Authority und einen als Hub. Ein guter Hub verlinkt viele wichtige Seiten mit hohem Authority Wert, eine gute Authority ist eine Seite, die von guten Hubs verlinkt wird. Hier als Analog zu PageRank wird Internet als ein Graph betrachtet. Man reduziert die Betrachtung auf einen Subgraphen, der relativ klein und reich an relevanten Seiten sein, sowie viele der stärksten Authorities enthalten sollte. Um so einen Subgraphen zu konstruieren, nimmt man eine Menge R besten Treffer einer textbasierten Suche, diese Menge R erweitert man zu einer größeren Menge S von Seiten, die zusammen mit den Links zwischen all den Seitenaus R sehr wahrscheinlich einen brauchbaren Subgraphen ergeben.

HITS bietet die Möglichkeit, nach ähnlichen Seiten zu suchen (die besonders viele gleiche Nachfolger und Vorgänger besitzen). Auch einer größe Vorteil ist die Berechnung den zwei Arten von Ranking, je nach Anwendungsfall kann das eine oder andere nützlicher sein.
Nach dem Vergleich der Algorithmen kann man sagen beide Verfahren anfällig für Manipulationsversuche sind, was sich allerdings bei PageRank weniger stark auswirkt. Und sie sind anfällig für Abschweifungen vom eigentlichen Thema (PageRank funktioniert sogar völlig unabhängig von Suchanfragen).