

# P2P

Jan Berweiler  
Philipp Hotz

November 2015

## **Zusammenfassung**

mein Kommilitone Jan Berweiler und ich nehmen in diesem Semester an der P2P-Vorlesung statt. Wir haben uns fuer das Semesterprojekt zu einer Gruppe zusammengeschlossen und da wir bisher noch keinen Vorschlag eingereicht haben, wuerden wir das hiermit gerne machen.

Wir haben uns dazu entschlossen eine verteilte Web-Suchmaschine zu schreiben, die sich aeusserlich dem Benutzer im Prinzip so darbietet wie auch Google und Konsorten, sich aber dahingehend von Ihnen absetzt, dass sie keine zentral gesteuerten Crawler fuer den Aufbau des Index benutzt, sondern die Benutzer des Programms nach und nach durch ihr eigenes Web-Surfing einen verteilten Index aufbauen, der dann ueber ein P2P-Overlay von allen genutzt werden kann.

Unser Programm wuerde dazu den HTTP-Traffic auf dem Rechner des Benutzers ueberwachen und eingehende HTML-Dokumente zusammen mit der URL in einen lokalen Index einfuegen. Ueber ein einfaches Interface soll der Benutzer Suchanfragen stellen koennen, die dann ueber ein Overlay an andere Peers verbreitet werden. Jeder Peer, der die Suchanfrage erhaelt, schaut dann in seinem Index nach, ob es relevante Dokumente (also in den meisten Faellen Webpages) gibt, und sendet gegebenenfalls diese als Liste zurueck. Alle auf diese Weise erhaltenen Suchergebnisse werden dem urspruenglichen Benutzer mit einem Ranking angezeigt, wobei Duplikate zuvor entfernt werden. Wie der Benutzer es von den bekannten Suchmaschinen gewohnt ist, soll ein Klick auf ein Suchergebniss dann zum jeweiligen Dokument fuehren.

Bisher saehe unser Ansatz so aus: um eingehende HTML-Dokumente in den Index einfuegen zu koennen, wuerden wir einen lokalen HTTP-Proxy benutzen. Dieser wuerde gleichzeitig dem Benutzer auch ein einfaches Webinterface bereitstellen, ueber das der Benutzer Suchanfragen stellen und Ergebnisse einsehen kann. Fuer den P2P-Teil wuerden

wir uns gerne eng am Vorbild von Gnutella orientieren. Wir wuerden es zu Beginn gerne so einfach wie moeglich halten, falls aber zum Schluss noch genuegend Zeit bleibt, auch Verfeinerungen wie Dynamic Querying einbauen.

## **1 P2P Search**