

UNIVERSITÄT LEIPZIG

INFORMATION RETRIEVAL LABORPROJEKT

Evaluation der Geocache-Suchmaschine

Steven Lehmann, Fabian Ziegner, Christian Schlecht

supervised by
Jun.-Prof. Dr. Martin POTTAST

18. September 2018

Inhaltsverzeichnis

1	Begriffserklärungen	2
1.1	Geocaching	2
2	Motivation	2
2.1	Domäne	3
2.1.1	Geocaching als Sport	3
2.1.2	technischer Stand	3
2.2	Dokumentkorporus	3
2.2.1	Dokumentstruktur	4
3	Architektur	4
4	Evaluation	5

1 Begriffserklärungen

1.1 Geocaching

Unter Geocaching wird eine Art Schnitzeljagd verstanden. Das Wort leitet sich vom lateinischen $\gamma\tilde{\eta}$, $g\tilde{e}$, stehend für "Erde" und dem englischen "cache" für "geheimes Lager" ab. Ziel ist es, mit Hilfe von GPS-Koordinaten und/oder veröffentlichten Tipps einen versteckten Behälter - den sogenannten "Geocache" (auch kurz "Cache") - zu finden. In der Regel beinhaltet dieser Behälter ein Logbuch, in welches sich die Geocacher bei erfolgreicher Suche eintragen können. Zusätzlich, sofern es die Größe des Behälters hergibt, sind oftmals kleine Gegenständen zum Tausch vorhanden. Wichtig ist es, während der gesamten Zeit möglichst unentdeckt zu bleiben, sodass der Cache stets verborgen bleibt. Nach Beendigung der Suche wird der Cache wieder an seine ursprüngliche Position gebracht.

Es gibt eine ganze Reihe Arten von Geocaches. Vom einfachen traditionelle Caches, über Multi-Caches und Rätsel-Caches bis hin zu virtuellen Caches können Suchen veranstaltet werden.

Fairness wird beim Geocaching groß geschrieben. Personen, die den Cache verändern, mutmaßlich zerstören, entfernen oder Dinge daraus stehlen wird größter Missmut entgegen gestellt.

2 Motivation

2.1 Domäne

2.1.1 Geocaching als Sport

Geocaching erfreut sich zunehmender Beliebtheit unter der Bevölkerung, obwohl es in gewissem Sinne das Ziel ist, geheim zu agieren. Dieser Sport kann von Menschen aller Altersgruppen und körperlicher Verfassung betrieben werden, da in praktisch jeder Region (Wald, Stadt, Parkanlagen, Gebirge, ...) Caches versteckt und gefunden werden können. Unterbewusst wird dadurch die Gesundheit gefördert. Aus dieser Hinsicht ist der Sport in jedem Fall zu unterstützen.

2.1.2 technischer Stand

Mittlerweile wird Geocaching durch das Internet stark unterstützt und vereinfacht. Während in den Anfängen (1980er Jahre) die Hinweise und Koordinaten meist verbal oder handschriftlich verbreitet wurden, gibt es heutzutage diverse Plattformen und Foren, um sich auszutauschen. Zu den bekanntesten gehören hier geocaching.com und open caching.de im deutschsprachigen Raum. Diese Seiten agieren im Wesentlichen als Foren zum Austausch mit anderen Cachern. Sie bieten zwar die Möglichkeit der Suche, jedoch wird hierbei auf eine Datenbank zurückgegriffen, aus welcher mittels SQL-Anweisungen Suchanfragen bearbeitet werden. Wir haben keinen Anbieter feststellen können, der bei der Suche Methoden des Information Retrieval, d.h. Verwendung eines Index und Retrieval-Modells, anwendet. An dieser Stelle wird im Rahmen dieses Projekts prototypisch gearbeitet.

2.2 Dokumentkorporus

Die Geocaching Suchmaschine soll für Geocaches in Deutschland angewendet werden. Aus diesem Grund wurde als Dokumentkorporus eine Sammlung von allen ¹ Geocaches in Deutschland, welche auf open caching.de veröffentlicht sind, gewählt. Der Korpus beläuft sich damit auf eine Größe von ca. 20.000 Caches. Aufgrund des Mehraufwandes eines Web-Crawls wurden diese manuell heruntergeladen. Die Dokumentsammlung kann also als statisch betrachtet werden.

¹Stand 24.04.18

```

Name: Unverhofftes Glück von Beeped Piano
Koordinaten: E 009° 27.066' N 49° 05.045'
Status: kann gesucht werden
Zustand: ok

Versteckt am: 01.11.2003
Wegpunkt: OC0ABC
Land: Deutschland
Cacheart: normaler Cache
Behälter: normal
D/T: 3.5/4.5
Online: https://www.opencaching.de/viewcache.php?wp=OC0ABC

Kurzbeschreibung: Am besten parkt man bei N 49° 05.077 E 009° 27.161

Beschreibung:
<=====
Hallo Freunde!
Habe heute einen 2. Cache versteckt.
Vorab, das Gelände ist ein bisschen schwierig. Die Kooords weichen unterschiedlich ab, mal 6 Meter mal 13Meter.
Zur Geschichte der Gegend erfährt man auf einer Hinweistafel am Parkplatz mehr.
Der Silberstollen ist nicht begehbar,dafür aber das Soldatenglück. Wer interesse hat hineinzugehen,sollte Gummistiefel und Taschenlampe nicht
So nun viel Spass am suchen.

Cacheinhalt: 1 Stein der Weissen
             1 Spiel die Siedler 2
             1 Regen Poncho
             1 Taschenlampenuhr
             1 lustiger Schlüsselanhänger
             2 mal eine Kinder Überraschung(nichts essbares)
             1Kugelschreiber sowie ein Bleistift
             1 Taschenrechner (solar)
und das wichtigste: 1 Log Buch

© Beeped Piano, Opencaching.de, CC BY-NC-ND, Stand: 24.04.2018;
alle Logeinträge © jeweiliger Autor
<=====

Zusätzliche Hinweise:
<=====

<=====
A|B|C|D|E|F|G|H|I|J|K|L|M
N|O|P|Q|R|S|T|U|V|W|X|Y|Z

Logeinträge:
<=====
A Meise / 06.04.2018 20:35 / gefunden

```

Abbildung 2.1: Dokumentstruktur eines Geocaches

2.2.1 Dokumentstruktur

Jedes Dokument kann von opencaching.de in diversen verschiedenen Formaten heruntergeladen werden, darunter auch direkt von GPS-System verwertbare Formate. Aufgrund der Vollständigkeit der Information wurden in diesem Rahmen Caches im .txt-Format verwendet.

Das Dokument wird bereits beim Herunterladen in eine wohlgeformte Struktur gebracht. Die Abbildung 2.1 zeigt einen Ausschnitt. Es lässt sich eine Art Key-Value-Struktur erkennen, welche das spätere Parsing der Dokumente vereinfacht, sodass eine maximale Korrektheit gewährleistet werden kann. Interessant ist hierbei, dass durch dieses Dokument sämtliche Informationen aus der Web-Darstellung übernommen werden, die Informationen sind also a priori vollständig.

3 Architektur

tbc

4 Evaluation

tbc