

Entwicklung und Evaluation eines multimodalen Empfehlungssystems für Lokationen



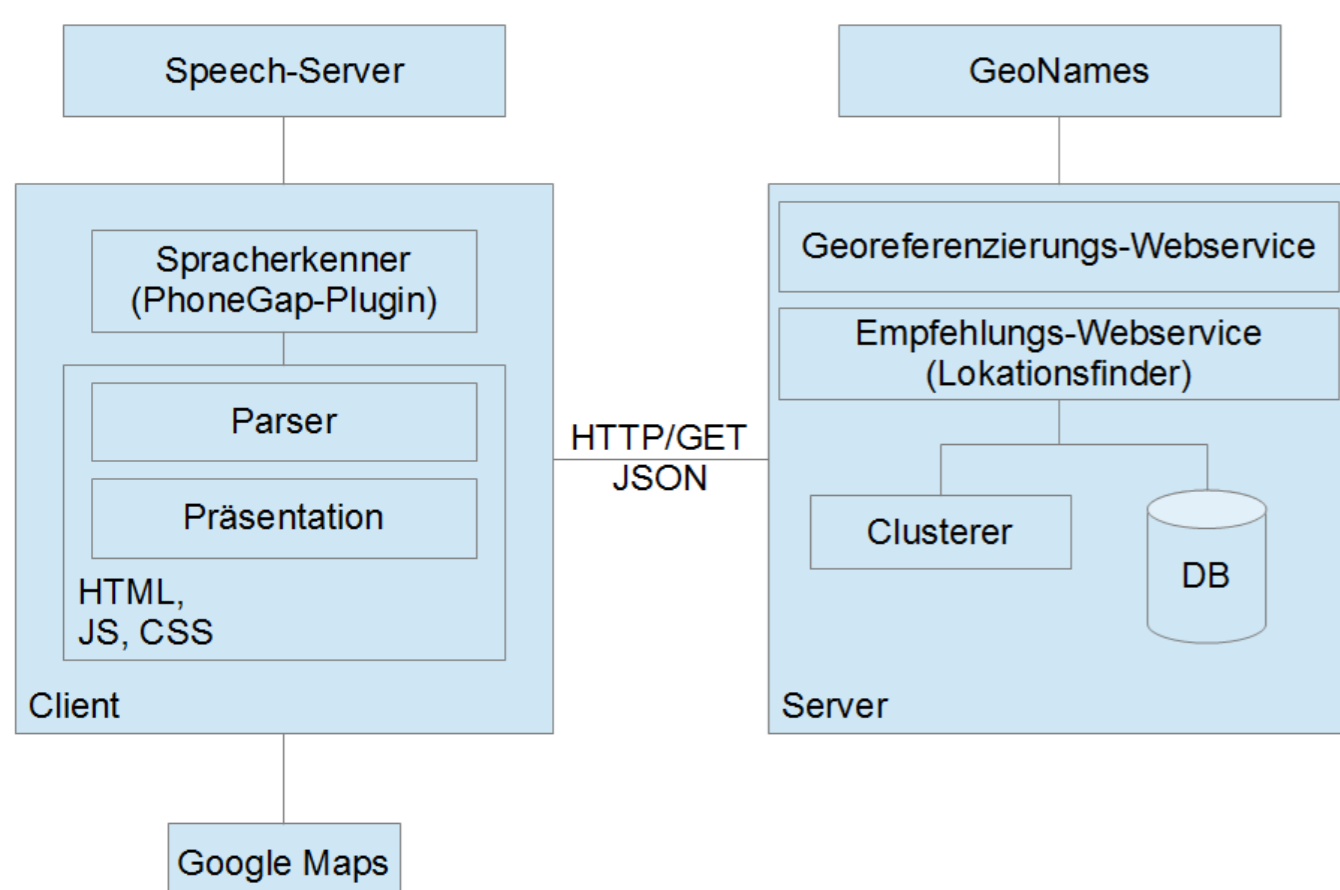
Bachelorarbeit, vorgelegt von Benjamin Hoffmann

Aufgabenstellung:

Zielstellung des Themas ist die Entwicklung und anschließende Evaluation eines multimodalen Empfehlungssystems für Lokationen. Es soll ein Prototyp einer App für Android-Smartphones entwickelt werden, die es ermöglicht, Orte unter Angabe einer Aktivität („essen gehen“) und einer Region („Berlin Mitte“) zu finden.

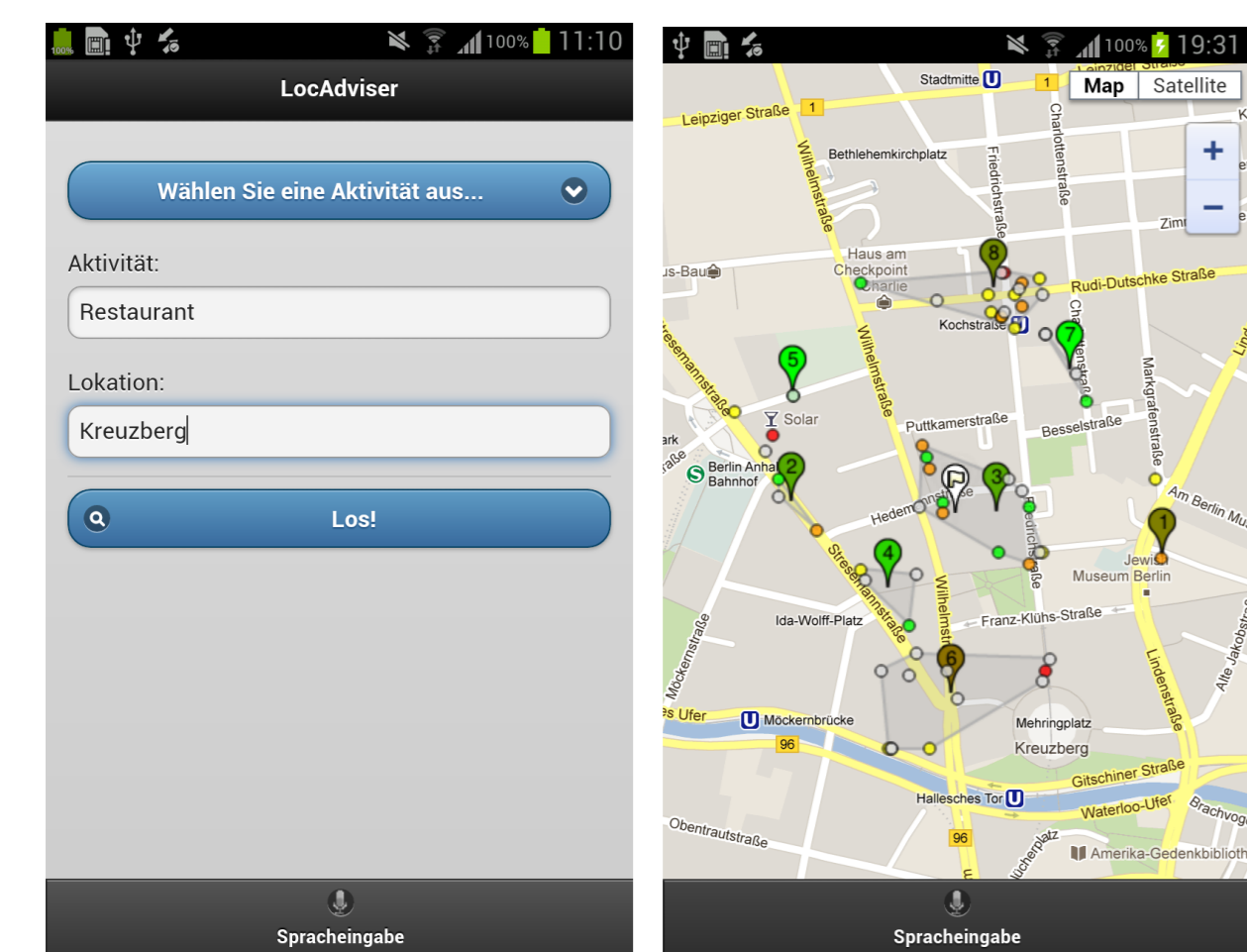
Die Arbeit konzentriert sich auf die Beantwortung zweier Hypothesen. **H1:** Für die Erledigung von Aufgaben in urbanen Gebieten (Einkaufen, Freizeitgestaltung, et cetera) können Daten aus sozialen Netzwerken genutzt werden, um den Benutzer bei der Auswahl geeigneter Lokalitäten zu unterstützen. **H2:** Die Bedienung einer Smartphone-Applikation zum Finden von Lokationen kann durch die Modalität Sprache erleichtert werden.

Systemarchitektur:



Das System verwendet eine Client-Server-Architektur. Der Einsatz des PhoneGap-Frameworks ermöglicht eine schnelle Entwicklung des Clienten mit Hilfe von Webtechnologien und vereinfacht das Portieren auf andere Plattformen. Der Client übernimmt insbesondere die Steuerung der Eingabe und die Präsentation der Ergebnisse.

Der Server stellt zwei RESTful-Webservices zur Verfügung: Einer ist für die Georeferenzierung zuständig, das heißt die Umwandlung von einem Ortsnamen zu den entsprechenden geografischen Koordinaten. Dazu wird auf den Dienst von GeoNames zugegriffen. Der zweite Webservice bestimmt geeignete Lokationen und liefert diese gruppiert zurück.



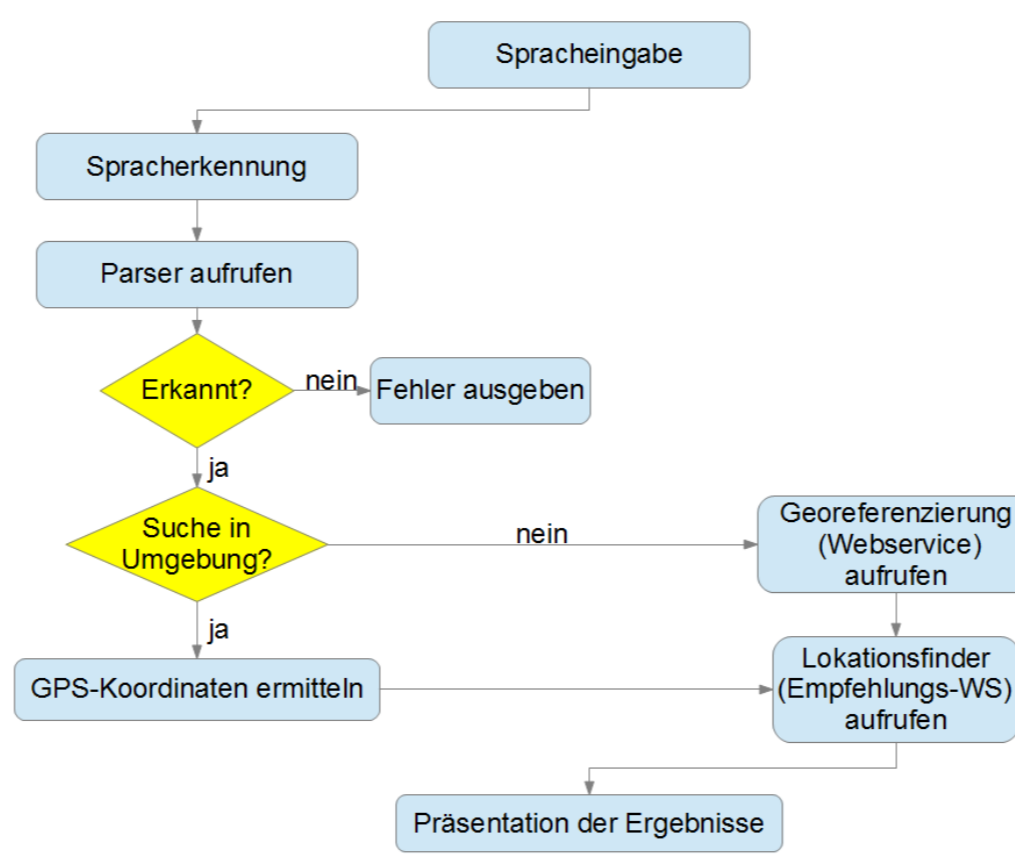
Benutzeroberfläche der Anwendung

Sprachsteuerung:

Die Eingabe kann sowohl via Touch als auch via Sprache durchgeführt werden. Die Definition einer Grammatik erlaubt die Beantwortung verschiedener Variationen. Beispiele für die initiale Suchanfrage:

- „Zeige mir bitte mal Restaurants in Berlin Kreuzberg“
- „Ich möchte am Hauptbahnhof essen gehen“
- „Wo kann ich Kaffee trinken am Alexanderplatz“
- „Suche bitte Pizzerien in der Nähe“

Sprachbefehle innerhalb der Kartenansicht erlauben das Wechseln der Ansichten, das Anzeigen der POIs innerhalb der einzelnen Cluster und das Verändern der Suchparameter.



Schematisierter Ablauf der Spracheingabe

Evaluation:

Eine Evaluierung der entwickelten App wurde mit 12 deutschsprachigen Mitarbeitern des Deutschen Forschungszentrums für Künstliche Intelligenz durchgeführt.

Die Versuchsteilnehmer mussten jeweils sechs Aufgaben lösen. Bei den ersten vier Aufgaben wurde die für die initiale Eingabe zu benutzende Modalität vorgeschrieben.

Das Experiment fand in einer kontrollierten Umgebung statt und wurde mit einer Videokamera für die spätere Auswertung aufgezeichnet. Nach dem Lösen der Aufgaben mussten die Probanden Fragen zu der Anwendung beantworten.

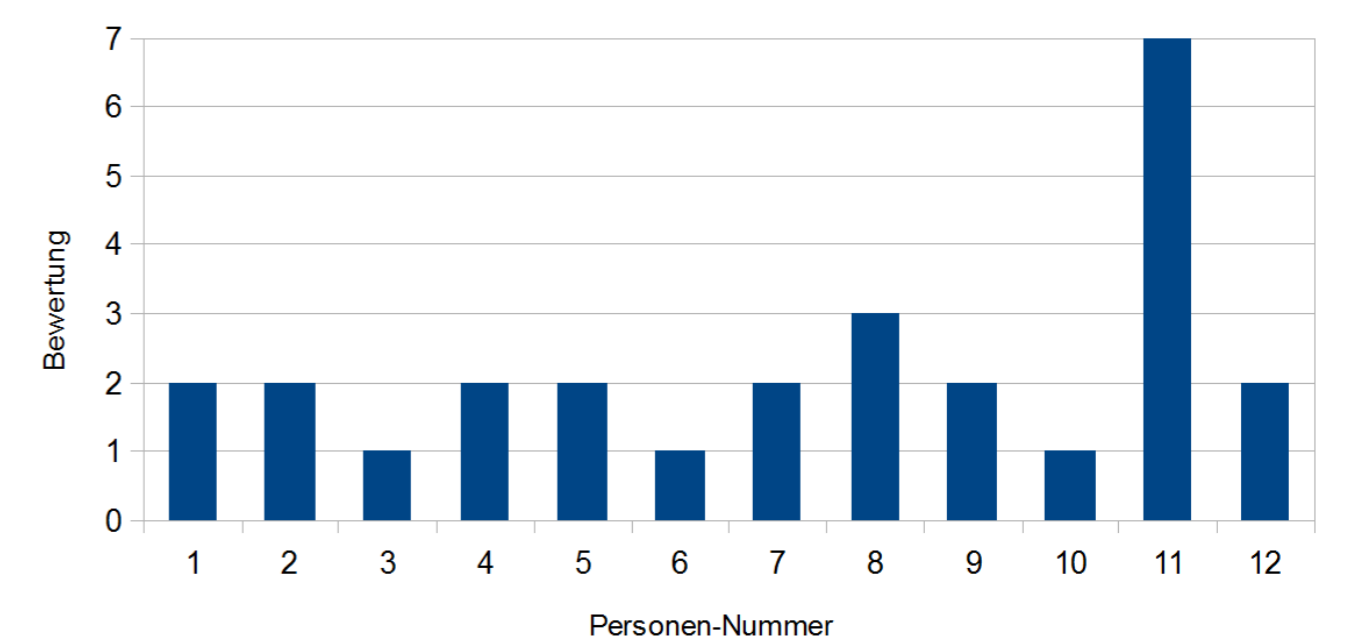
Aufgaben-Nr.	Aktivität	Ort
1	Essen	Berlin Friedrichshain
2	Café	Berlin Hauptbahnhof
3	Einkaufen	Berlin Alexanderplatz
4	Sushi	Berlin Mitte
5	Pizzeria	Berlin Tempelhof
6	Am Abend etwas unternehmen	<offen>

Auswertung:

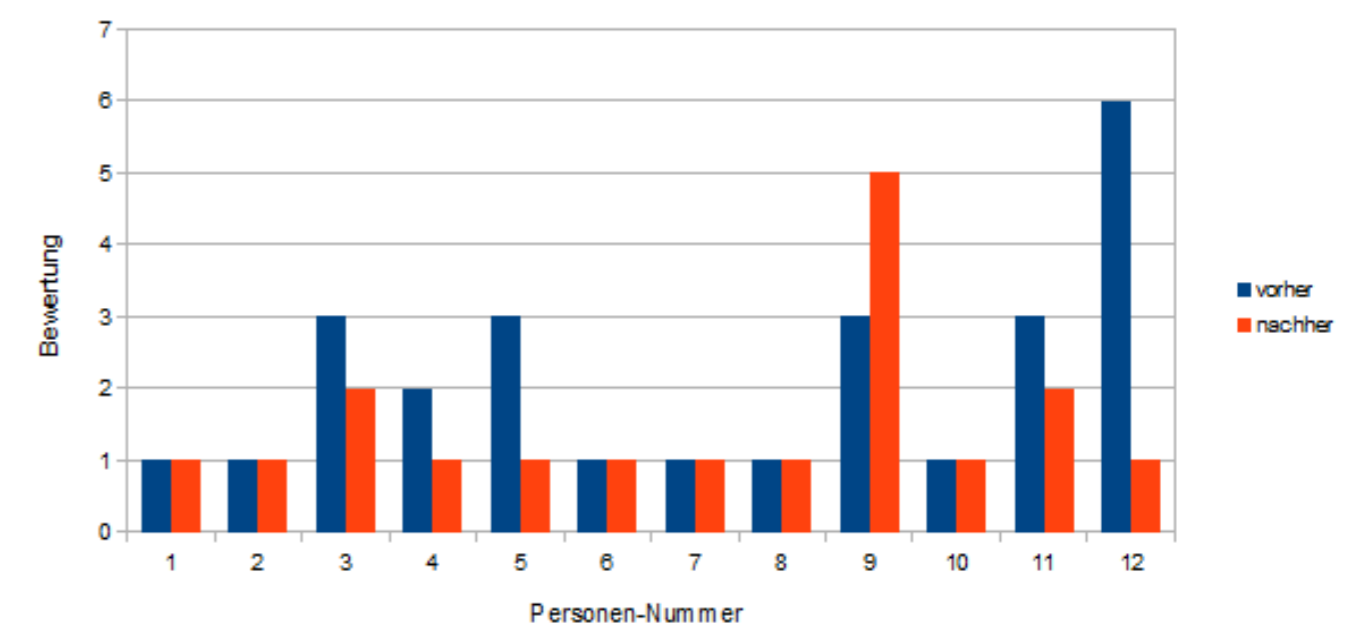
Bei den zwölf Teilnehmern (9 männlich, 3 weiblich) lag der Median des Alters bei 33,5 Jahren (Durchschnitt: 36,25; Minimum: 26; Maximum: 53). Ein Großteil von ihnen arbeitet als Forscher im Bereich Informatik beziehungsweise Linguistik. Sieben Personen (58,34%) gaben an, ein Smartphone zu besitzen.

Im Vorab-Fragebogen wurde nach der Erfahrung mit Sprache als Eingabemodalität gefragt. Ungefähr die Hälfte (7 Probanden) gab an, dass sie Erfahrung mit Spracheingabe hatten.

Um die erste Hypothese zu überprüfen, wurden die Probanden gefragt, ob die Informationen aus sozialen Netzwerken hilfreich waren. Elf der 12 Befragten sagen aus, dass sie dieser Aussage eher zustimmen (auf einer Likert-Skala von 1 [stimme zu] bis 7 [stimme nicht zu]; Durchschnitt: 2,25; Median: 2). Nur eine Person antwortete, dass Empfehlungen nicht hilfreich seien.



Die Einschätzung von Sprache als Eingabemodalität wurde vor und nach dem Lösen der Aufgaben von den Probanden vorgenommen. Die Hälfte der Versuchsteilnehmer schätzten sie konstant für sehr gut ein. Fünf Personen bewerteten die Nützlichkeit nach dem Experiment besser. Nur eine Person korrigierte ihre Einschätzung zum Schlechteren. Ein Proband (Nummer 12) änderte sogar komplett seine Meinung – von einer sehr negativen Einstellung zu einer sehr positiven.



Bei der fünften und sechsten Aufgabe entschied sich die Mehrheit für die Modalität Sprache: 16 der 23 Aufgaben (69,5%) wurden mittels Spracheingabe gelöst, 4 mittels Toucheingabe und 3 via Kombination aus Touch- und Spracheingabe.

Die Videoauswertung ermöglichte eine objektive Analyse der beiden Modalitäten. Bei der Auswertung der Zeiten wurden jeweils die ersten vier Aufgaben betrachtet. Die durchschnittliche Zeit einer Spracheingabe mit entsprechender Antwort dauert 18,51 Sekunden (Median: 13 Sekunden; Minimum: 9,11 Sekunden; Maximum: 56,78 Sekunden). Die durchschnittliche Dauer bei der Bedienung mittels Touch beträgt 45,30 Sekunden (Median: 43 Sekunden; Minimum: 24,3 Sekunden; Maximum: 1:32,19). Diese Ergebnisse zeigen, dass die Eingabe via Sprache im Durchschnitt mehr als doppelt so schnell war.

Fazit:

In dieser Arbeit wurde die prototypische Entwicklung und Evaluation eines multimodalen Empfehlungssystems für Lokationen beschrieben. Leichte Modifizierungen vorhandener Komponenten – wie Spracherkennung und Grammatikparser – wurden erfolgreich mit Diensten aus dem Internet zu einer multimodalen Anwendung kombiniert.

Die beiden Hypothesen, dass Daten sozialer Netzwerke bei der Entscheidungsfindung helfen (H1) und Sprache die Eingabe erleichtert (H2), konnten durch die Evaluierung bestätigt werden.