

UNIQUES - Partnervorschläge auf Partnervermittlungsplattform

Studiengang: BSc in Informatik | Vertiefung: Computer Perception and Virtual Reality

Betreuer: Prof. Dr. Bernhard Anrig

Experte: Han van der Kleij (SBB)

Die Liebe des Lebens finden, doch wie? Immer mehr Menschen suchen ebendiese über Online-Partnerplattformen. In der unübersichtlichen Masse an Benutzerprofilen den richtigen Partnern zu finden gestaltet sich zuweilen als grosse Herausforderung für den Benutzer. Ziel dieser Bachelorarbeit ist das Verfassen eines Algorithmus, welcher einen Benutzer bei dieser Suche unterstützt.

1

UNIQUES

Als Datenquelle für den Algorithmus wurde die sich nach wie vor in der Entwicklungsphase befindende Online-Partnerplattform UNIQUES verwendet. Diese unterstützt sowohl ein dynamisch konfigurierbares Set an Profilattributen als auch diverse Schnittstellen, um Daten zu beziehen und zu schreiben. Vorbereitend für die Bachelorarbeit wurde ein Tool umgesetzt, welches die Generierung von Beispiel-Profilen ermöglicht, da die Plattform zum aktuellen Zeitpunkt noch nicht aufgeschaltet ist und daher keine realen Daten zur Verfügung stehen.

Architektur

Eine primäre Anforderung an eine Website ist deren Antwortzeit; das Bewältigen von grossen Datenmengen ist jedoch sehr zeitaufwändig. Um einem Benutzer trotz den grossen zu bewältigenden Datenmengen zeitnah Vorschläge liefern zu können, wurde eine Lösung mit zwei Komponenten umgesetzt: eine sehr rechenintensive Clustering-Komponente und eine schnelle Vorschlags-Komponente. So können die rechenintensiven Operationen von der Benutzerinteraktion getrennt und die Architektur in der Folge beiden Anforderungen gerecht werden.

Clustering

Die von UNIQUES bezogenen Profildaten werden mittels einer kombinierten Metrik, welche die unterschiedlichen Attribute messen kann, auf Ähnlichkeiten überprüft. Basierend auf gängigen Clustering Algorithmen wird eine dynamische Anzahl von Clustern erzeugt. Dabei wird mittels analytischen Möglichkeiten versucht, eine ideale Clusteranzahl zu bestimmen. Das Resultat wird basierend auf der besten Konfiguration von Clustern berechnet. Um mehr Flexibilität zu erhalten, wird die Clusterzugehörigkeit der einzelnen Profile in Fuzzy Form abgespeichert, sodass ein Profil nicht strikt zu einem Cluster gehört, sondern zu einer Kombination aus solchen.

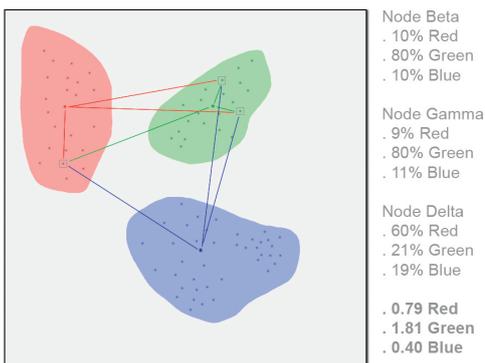
Eine besondere Herausforderung an den Algorithmus stellen die grossen Datenmengen dar. Ziel ist es, einen Datenraum von 100 000 Profile behandeln zu können. Hierfür müssen zahlreiche Berechnungen ausgeführt werden. Um dennoch in sinnvoller Zeit ein Resultat zu erhalten, wurden sowohl Resultat-Caching als auch Multi-Threading implementiert. Die erwartete Laufzeit liegt jedoch nach wie vor bei rund acht Stunden.

Vorschläge

Die Berechnung der Vorschläge für einen Benutzer basiert nicht auf einem psychologischen Ansatz, sondern auf der Auswahl des Benutzers. Der Benutzer wählt auf der Plattform Profile von Benutzern aus, welche ihm ge- bzw. missfallen. Aufbauend auf ebendieser Auswahl berechnet die Vorschlags-Komponente eine Liste mit Profil-Vorschlägen. Dabei können die vorher berechneten Informationen aus dem Clustering verwendet werden, um dem Benutzer in Sekundenschnelle eine Liste mit für ihn interessanten Profile zu präsentieren.



Christoph Seiler



Fuzzy Selektion von Profilen aus Clustern