Das Web als personalisierte Entscheidungsplattform – Die PESCaDO Idee

Harald Bosch, Dennis Thom, Thomas Ertl,

Institut für Visualisierung und Interaktive Systeme, Universität Stuttgart,

{Harald.Bosch, Dennis.Thom, Thomas.Ertl}@vis.uni-stuttgart.de

Abstract: Anwenderorientierte Datendienste im Web 2.0 wie Karten-, Wetter oder Verkehrsportale bieten ihren Nutzern in Alltagssituationen eine praktische Entscheidungshilfe. Dabei sind die Nutzer jedoch gezwungen diese Dienste selbst zu erschließen, auszuwerten und für ihr spezifisches Problem zu verknüpfen. Im PESCaDO-System sollen frei verfügbare Dienste mittels automatischer Extraktion als Grundlage für eine generische und integrierte Entscheidungsunterstützung im Kontext von Umweltdaten dienen. Den Nutzern werden dabei auf Basis eines intelligenten Anfrageeditors und einer semantischen Anfrageverarbeitung durch visuelle Repräsentationen relevante Informationen und Planungsvorschläge unterbreitet.

1 Einführung

Die Verfügbarkeit interaktiver Web 2.0-Dienste bedeutet für viele Menschen eine erhebliche Erleichterung von Planungs- und Entscheidungssituationen des alltäglichen Lebens. So können etwa kostenlose Karten- und Wetterdienste genutzt werden, um Ort und Zeit für eine geplante Freizeitaktivität zu bestimmen. Mittels Verkehrsinformations- und Gesundheitsportalen lässt sich dann sicherstellen, dass das Ausflugsvorhaben nicht durch eine drohende Autobahnsperrung oder eine hohe Belastung der Gegend durch Allergieerreger zum Debakel gerät. Nichtsdestotrotz sind die Nutzer solcher Dienste gezwungen, die jeweils passenden Angebote selbst zu ermitteln und ihre Anfragen entlang der Möglichkeiten eines Dienstes zu orientieren und spezifisch zu formulieren. Weiterhin müssen die angebotenen Informationen der verschiedenen Dienste bezogen auf den Problemkontext manuell ausgewertet, miteinander verknüpft und ihre Zuverlässigkeit und Eignung für die jeweilige Situation beurteilt werden, was aufgrund der Darbietung der Daten in vielen unterschiedlichen Formaten und Visualisierungsformen häufig mit einem erheblichen Aufwand verbunden ist.

Aufgrund dieser Schwierigkeiten erfreuen sich sogenannte Web-Mashups einer immer größeren Popularität. Dabei handelt es sich um collageartige Kombinationen unterschiedlicher Web 2.0-Dienste, welche neben den interaktiven Webseiten für Endanwender auch zumeist REST-basierte Programmierschnittstellen (API) für den Zugriff auf ihre Daten zur Verfügung stellen. Obschon die Nutzer auf diesem Weg zwar viele relevante Infor-

erschienen im Tagungsband der INFORMATIK 2011
Lecture Notes in Informatics, Band P192
ISBN 978-3-88579-286-4
weitere Artikel online:
http://informatik2011.de/519.html

mationen in bereits verknüpfter Weise abrufen können, handelt es sich auch hier nur um problemspezifische Lösungen, bei denen die Möglichkeiten der Interaktion mit den angebotenen Daten und die Konfigurierbarkeit und Auswahl der visuellen Repräsentation nur in sehr begrenztem Umfang möglich sind. Auch hat der Nutzer hier zumeist keine Möglichkeit, die Herkunft und Zuverlässigkeit der jeweiligen Daten zu prüfen und ggf. einzelne Datenquellen durch vom System vorgeschlagene Alternativen zu ersetzen, denen er mehr Vertrauen entgegenbringt.

Das EU-geförderte Projekt PESCaDO (Personalized Environmental Service Configuration and Delivery Orchestration) plant ein System, das diesen Problemen begegnen soll. Entwickelt wird hierbei eine personalisierte Entscheidungsunterstützung auf Basis von automatisch extrahierten Informationen aus dem Web. Da sich bei den Bürgern der Europäischen Union ein zunehmendes Bewusstsein für den Einfluss von Umweltbedingungen auf ihre persönliche Lebensqualität beobachten lässt, liegt bei PESCaDO die Berücksichtigung von Umweltinformationen – d.h. meteorologischen Informationen, Luftqualität, Pollenbelastung und Verkehr – im Fokus. Das entstehende System unterstützt den Anwender von der Formulierung seines Problems bzw. Informationsbedarfs, über die Auswahl der Datenbasis bis hin zur Interpretation der Ergebnisse.

In dieser Arbeit soll die Frage erörtert werden, wie man die Komplexität einer semiautomatischen, personalisierten Entscheidungsunterstützung vor dem Benutzer verbergen und den Aufwand des Nutzers auf erwünschte und auf den jeweiligen Problemkontext angepasste Interaktionsmöglichkeiten reduzieren kann. Zu diesem Zweck müssen Abstraktionsebenen der Interaktion gefunden werden, welche zum einen eine generische Verarbeitung von Daten aus verschiedensten Web-Quellen ermöglichen und zum anderen die Instanziierung von effizienten und einfach bedienbaren Interaktionselementen erlauben. Die generische Konfigurierbarkeit, Vergleichbarkeit und Orchestrierung der Darstellung verfügbarer Information soll im vorgestellten Ansatz durch eine semantische Zwischenschicht in Form einer Umweltdaten-Ontologie erreicht werden, welche das Anfrage- und Visualisierungssystem des Anwenders von den verfügbaren Rohdaten und Services im Web trennt und zwischen ihnen vermittelt. Ein intelligentes Anfragesystem soll ferner die gezielte und auf den persönlichen Hintergrund und gewählten Problemkontext angepasste Steuerung der Problemverarbeitung ermöglichen. Dies geschieht auf Basis einer mit der Umwelt-Ontologie verknüpften Problem-Ontologie und einer eigens entwickelten Problembeschreibungssprache (*Problem Description Language*, kurz PDL) sowie der Verwendung zuvor erstellter Nutzerprofile, um eine personalisierte Beschreibung zu gewährleisten.

In den folgenden Kapiteln werden wir zunächst kurz die verwandten Arbeiten in diesem Bereich vorstellen, die Anforderungen und den Aufbau des geplanten Systems beschreiben und erläutern, welche Möglichkeiten daraus für die Interaktion und Visualisierung entstehen.

2 Verwandte Arbeiten

Konzeptionell basiert PESCaDO auf den Vorarbeiten des ebenfalls von der EU geförderten Projekts MARQUIS [WNBA+07] (Multimodal Air Quality Information Service for General Public). MARQUIS erstellt automatisch an die Nutzergruppe angepasste Berichte über die Luftqualität in ausgewählten Regionen. Hierbei lag der Forschungsschwerpunkt auf der Generierung natürlichsprachlicher Berichte in der jeweiligen Landessprache. Anders als PESCaDO arbeitet MARQUIS auf einer wohldefinierten und auf Luftqualität beschränkten Datenquelle. Es erlaubt nur eine Personalisierung anhand existierender Profile für Nutzergruppen wie z.B. „Öffentlichkeit“, „medizinischer Spezialist“ oder „Sportler“ und unterstützt keine interaktive Erstellung komplexerer Anfragen.

Ein weiteres Projekt, welches sich mit der Entscheidungsunterstützung aufgrund von Umweltinformationen beschäftigte, ist das ECOSIM Projekt[1]. ECOSIM ist ein model-basiertes Informationssystem das überwiegend im städtischen Umweltmanagement eingesetzt werden kann. Es erlaubt die Erstellung und Simulation numerischer Modelle und kann damit beispielsweise die Auswirkungen von Bauvorhaben oder die Ausbreitungen von Verschmutzungen prognostizieren. ECOSIM ist ein Expertensystem, bei dem die Benutzer direkt an der Datenbasis arbeiten und ein entsprechendes Hintergrundwissen benötigen. In PESCaDO soll diese Komplexität in den semantischen Ressourcen enthalten sein damit der Benutzer sich in seinem Problemkontext bewegen kann, ohne auf die dafür benötigte Datenbasis achten zu müssen.

Um dies zu erreichen, soll die Anfrageerstellung möglichst uneingeschränkt und direkt, aber dennoch jederzeit nachvollziehbar und intelligent unterstützt werden. Einen guten Überblick über die Historie auf dem Gebiet der visuellen Anfrageerstellung bietet hier das Paper von Catarci [CCLB97]. Bisherige visuelle Anfragesysteme sind meist auf bestimmte Datenformate oder Anfragesprachen (z.B. XML oder SQL) zugeschnitten. Die Arbeit von Dongilli et al. [DFT04] im Projekt SEWASIE[2] beschreibt eine Möglichkeit wie Anfragen an heterogene Datenbasen von einem Benutzer erstellt werden können, dem das Vokabular der Daten nicht bekannt ist. Dazu wird eine die Datenbasis beschreibende Ontologie eingesetzt, welche näher an der Ausdrucksweise des Benutzers liegen soll und ihm damit als Einstiegspunkt dienen kann. Im Zusammenspiel mit natürlichsprachlichen und visuellen Darstellungen der Anfrage soll der Benutzer im Laufe der Interaktion das Vokabular der Datenbasis erlernen. PESCaDO setzt ebenfalls eine Ontologie als Abstraktionsebene zwischen Anfrage und Datenbasis ein, jedoch dient diese auch als Grundlage der Auswertung der Anfragen.

Als verwandte Arbeit im Bereich der dynamischen Integration von Anfrage und Datenbasis sind auch Dong und Loo [DL01] zu nennen. Sie beschreiben die Architektur eines Systems das neue webbasierte Decision Support Systeme (DSS) generiert. Hierbei sollen Software-Agenten eingesetzt werden, um auf der Basis von Szenarien die Komponenten für das DSS flexibel zusammenzuführen. Die dabei verwendeten Komponenten umfassen: User Interface, Data, Model, Solver und Visualization. Leider gibt es keine aktuelleren Informationen zum Fortschritt dieses Vorhabens.

[1] http://www.ess.co.at/ECOSIM/
[2] http://www.sewasie.org/

3 Konzeption von PESCaDO

In PESCaDO sollen verschiedenste Webseiten und Dienste genutzt werden, um dem Nutzer eine personalisierte Entscheidungsgrundlage zur Verfügung zu stellen. Um dabei nicht vollständig auf bestehende Webservices von Datenanbietern angewiesen zu sein und um nicht nur weitere spezialisierte Mashups für konkrete Problemstellungen zu generieren, basiert die Entscheidungsunterstützung in PESCaDO auf der Auswertung semantischer Modelle. D.h. der Informationsbedarf des Nutzers wird unabhängig von der eigentlichen Datenbasis formuliert und dann im Kontext der Daten ausgewertet. Dafür existiert eine zentrale Ontologie um den Informationsbedarf des Anwenders und das Informationsangebot des Internets auf eine gemeinsame Basis zu bringen. Diese ist konzeptionell in drei Bereiche unterteilt. (1) Die Problembeschreibungssprache (PDL) teilt einerseits der Benutzungsoberfläche die möglichen Ausprägungen von Aktivitäten und Anfragearten mit, und stellt andererseits das Vokabular in dem die Anfragen an das System übermittelt werden. Die PDL ist unabhängig von der Domäne der Daten. (2) Die Beschreibung der Datendomäne wiederum fungiert als Basis in welche alle im Internet gefundenen Informationen übersetzt werden müssen. Im Fall von PESCaDO handelt es sich hierbei um Formate für meteorologische, luftqualitätsbezogene und pollenbezogene Informationen. (3) Die Beschreibung der Zusammenhänge inwiefern die Aktivitäten Anforderungen an die Daten stellen bzw. wie die Daten die Aktivitäten beeinflussen könnten.

Als Beispiel kann die Anfangs erwähnte Planung einer Freizeitaktivität dienen. Der Nutzer formuliert seine geplante Aktivität und seinen gesundheitlichen Zustand und überlässt es dem System daraus die notwendige Datenbasis abzuleiten und entsprechende Informationen aus dem Internet zusammenzuführen und zu verknüpfen.

Dieser Prozess der Entscheidungsunterstützung wird im Folgenden in vier Phasen unterteilt (siehe Abbildung 1). Vor der eigentlichen Benutzung des Systems werden mögliche Datenquellen für verschiedene Umweltaspekte (Wetter, Pollen, ...) und geographische Regionen gesucht, verarbeitet und mit weiteren Metadaten in einer Datenbank von Webseiten und Webservices (*Environmental Node Repository*) hinterlegt (Schritt 0 in Abb. 1). Während einer konkreten Sitzung übermittelt ein Anwender unter Berücksichtigung seines Profils, seines Kontextes und der Möglichkeiten der Problembeschreibungssprache (PDL) eine Anfrage an das System (Schritte 1-3). Anschließend wird die Anfrage mit den Informationen des *Node Repositories* verknüpft und eine Antwort wird erstellt (Schritte 4-6). Schließlich kann der Anwender das Resultat analysieren und gegebenenfalls Korrekturen an der Eingabe und den Verarbeitungsparametern vornehmen (Schritte 7+8). Die folgenden Unterkapitel beschreiben jeweils eine dieser Phasen.

3.1 Datengewinnung

Im Internet lässt sich heute eine Vielzahl umwelt- und umfeldbezogener Datenquellen finden, welche zumeist in Form von Webseiten oder Webservices verfügbar gemacht werden. Im Falle von Webservices, wie etwa der populären REST-APIs, lassen sich die verfügbaren

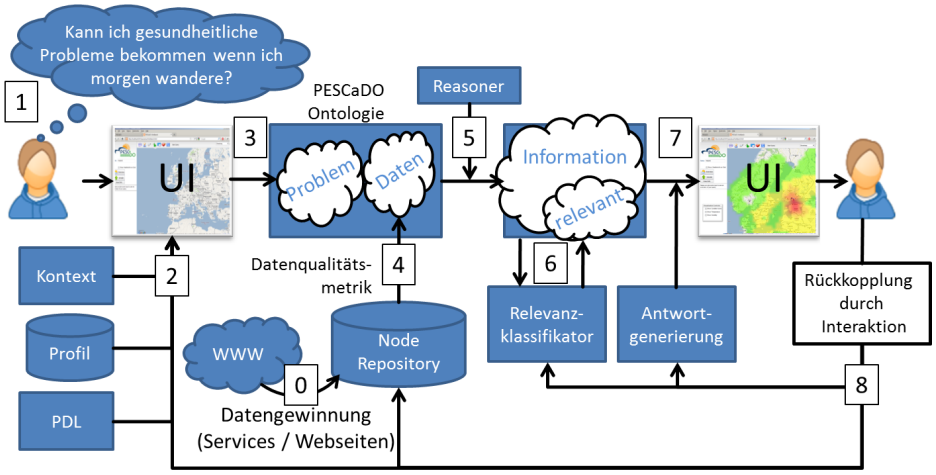

Abbildung 1: Die PESCaDO-Architektur: 1) Der Nutzer wendet sich mit seinem Problem an das System 2) Auf Basis von Problemkontext und Nutzerprofil wird eine Anfrage in PDL erzeugt und 3) in die Problemontologie überführt 4) Aus dem Node-Repository werden die relevanten Knoten ermittelt und die Daten in die Ontologie aufgenommen 5) Auf Basis der Anfrage erzeugt der Reasoner eine Antwortbasis 6) Der Relevanzklassifikator ermittelt die relevanten Informationen 7) In der Antwortgenerierung werden die Resultate in Form interaktiver und konfigurierbarer Visualisierungen präsentiert. 8) In einem iterativen Prozess kann der Nutzer auf die verschiedenen Elemente der Antwortgenerierung Einfluss nehmen.

Daten mittels definierter Schnittstellen häufig auf einfache Weise automatisiert und strukturiert extrahieren und für weitere Verarbeitungsschritte aufbereiten. Anders ist die Situation bei themenorientierten Webseiten (z.B. wetter.com oder verkehrsinfo.de), die keine Programmierschnittstelle anbieten, sondern die Daten in Form von Text oder Grafiken zur Verfügung stellen. In diesem Fall können bis zu einem gewissen Grad Metadaten (etwa XML-Annotationen), welche mit einzelnen Webseiten verknüpft sind, automatisch ausgewertet und etwa zur Klassifikation der Seiten und einzelner Unterseiten herangezogen werden; darüber hinaus sollen im PESCaDO-System aber auch Verfahren des Sprach- und Bildverstehens zur Anwendung kommen, um die relevanten Informationen aus Text und Grafiken zu extrahieren.

Zum Aufbau des *Environmental Node Repository* sollen neben einer vordefinierten Menge von bekannten Webseiten und Webservices auch neue und bisher unbekannte Datenknoten mit Hilfe von Websuchmaschinen aufgefunden, anschließend ausgewertet, indexiert und in das Node-Repository aufgenommen werden. Dies geschieht über eine domänenspezifische Websuche [Wöb06] durch die Verwendung von *keyword spices* [OKI04]. Für die jeweilige Domäne (z.B. Wetter, Luftqualität, Verkehr) spezifische Suchwörter werden unter Zuhilfenahme von Verfahren des maschinellen Lernens aus bekannten Webseiten der Domäne extrahiert und in eine neue Suchanfrage eingebracht.

Im Anschluss an die Suche werden die Webseiten nach ihrer Relevanz untersucht und nützliche von nutzlosen Ergebnissen separiert sowie eine Kategorisierung der verschiedenen Dienst- und Domänenarten vorgenommen. Ferner wird die Struktur der Seiten analy-

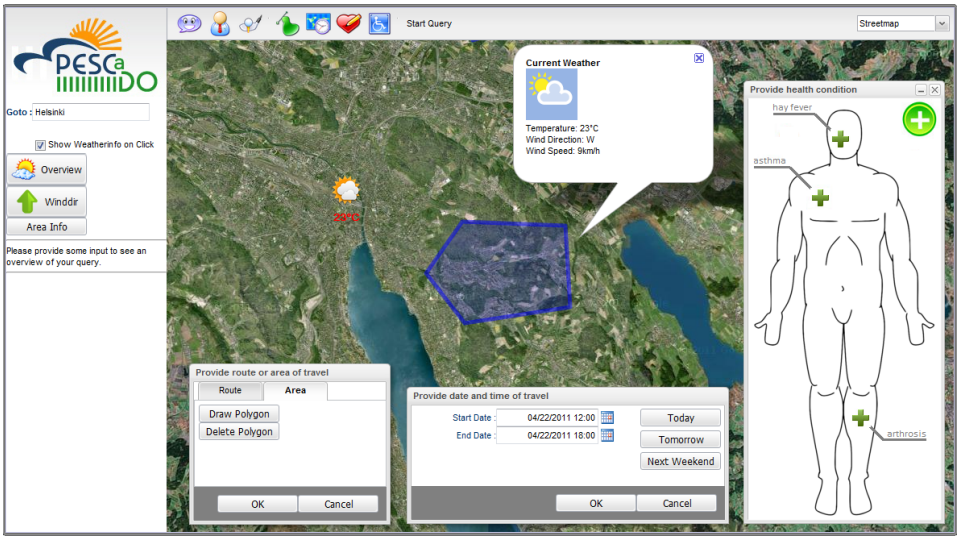

Abbildung 2: Screenshot der Anfrageformulierung mit verschiedenen Widgets. Zu sehen sind im mittleren Teil zwei Widgets zur Spezifikation von Zeitpunkt und Ort einer geplanten Aktivität. Rechts daneben befindet sich ein Widget, mit dem sich gesundheitliche Probleme des Nutzers angeben lassen.

siert und Zugriffsinformationen für die informationstragenden Elemente werden erzeugt. Nach der Analyse werden jeweils die Zugriffsadresse (URL) eines Service-Knotens sowie seine Klassifikation und weitere Metainformationen gespeichert. Mittels eines Crawlers sollen die bereits bekannten Knoten kontinuierlich nach Änderungen und neuen Daten oder Unterknoten untersucht und die Resultate gegebenenfalls zur weiteren Verwendung zwischengespeichert werden. Je nach Bedarf können die Datenknoten nun genutzt werden, um gezielt Informationen auszulesen, welche sich in die Strukturen der PESCaDO-Ontologie überführen und in der Wissensbasis repräsentieren lassen.

3.2 Anfrageformulierung

Die Interaktion des Nutzers mit dem PESCaDO-System soll verschiedene Anforderungen erfüllen. Eine möglichst generische aber zugleich einfache und leicht verständliche Bedienung soll durch eine intelligente Steuerung des Anfrageprozesses ermöglicht werden. Um einem weiten Feld an Nutzern hürdenlos zur Verfügung zu stehen, soll die Anfrage an das PESCaDO-System durch den Endbenutzer in einer webbasierten graphischen Benutzungsoberfläche erfolgen (Siehe Abb. 2).

Im Kern des Anfrageprozesses steht dabei eine aktuelle Problembeschreibung, welche im Hintergrund der UI-Interaktion auf Basis einer eigens entwickelten Problembeschreibungssprache (PDL) aufgebaut wird. Vor der eigentlichen Anfrage hat der Nutzer die Möglichkeit, sich die verschiedenen Anfrage- bzw. Problemtypen, die dem System auf

Basis der Problem-Ontologie bekannt sind, anzeigen zu lassen. Mit seiner Auswahl bestimmt er den aktuell gültigen Problemkontext – so etwa die Planung einer Wanderung unter Berücksichtigung besonderer gesundheitlicher Aspekte. Auf Basis dieses Kontextes und dem persönlichen Hintergrund des Nutzers in Form eines hinterlegten Nutzerprofils kann das System ableiten, welche Informationen für die Anfrage vom Nutzer mindestens angegeben werden müssen. Aus einer vordefinierten Bibliothek interaktiver Widgets und Anfragedialoge wird nun eine Menge an Werkzeugen ausgewählt, welche dem Benutzer zur Spezifikation seiner Anfrage angeboten werden. Beispielsweise kann ein graphisches Tool zur Erstellung einer Wanderroute auf einer Karte oder ein Kalender zur Angabe von gewünschter Start- und Endzeit präsentiert werden. Der Nutzer hat weiterhin die Möglichkeit, die Anfrage auf eigene Initiative hin um Aspekte zu ergänzen, die für das von ihm erwartete Resultat von besonderer Relevanz sind, wie etwa die Angabe von gesundheitlichen Gebrechen, welche im Nutzerprofil nicht erfasst sind oder die eine besondere Berücksichtigung bei der Anfrage erfahren sollen. Somit kann er seine Anfrage steuern und weitere Informationen in den Problemlösungsprozess einbringen, deren Relevanz aus Kontext und Hintergrund alleine nicht hervorgehen. Der gegenwärtige Stand der Problembeschreibung wird dem Nutzer unter Zuhilfenahme der präsentierten Visualisierungen (Route, Uhrzeit etc.) in Form eines Anfrageberichts angezeigt. Nachdem mindestens die minimalen Informationsanforderungen gemäß der Problem-Ontologie erfüllt sind, kann der Nutzer seine Anfrage an das System starten.

3.3 Anfrageverarbeitung

Ausgehend von der in der Anfrage enthaltenen Problembeschreibung, dem Nutzerkontext und dem Benutzerprofil können nun alle möglicherweise relevanten Umweltaspekte gesammelt und aus dem *Node Repository* mögliche Datenquellen identifiziert werden. Diese umfassen üblicherweise mehr Aspekte als für die Beantwortung des Anfrage notwendig wären, da z.B. eine körperliche belastende Aktivität im Freien automatisch zur Abfrage aktueller Ozonkonzentrationen führen würde, ungeachtet der aktuellen Jahreszeit. Aus den möglichen Datenquellen müssen anschließend diejenigen ausgewählt werden, welche am besten zum Problemkontext des Nutzers passen. Hierfür sind Konfidenz- und Datenqualitätsmetriken nötig um die Datenquellen quantitativ vergleichen zu können. Diese Metriken werden von den Administratoren des Systems gepflegt, können aber vom Endbenutzer für seine Sitzung in gewissem Maße verändert werden.

Nachdem die Datenquellen für die Anfrage feststehen, werden die eigentlichen Werte abgerufen. Mit Hilfe von Methoden der Sprachverarbeitung, und in Ausnahmen auch der Bilderkennung, wird hierbei versucht nur die relevante Information zu extrahieren und Inhalte der Seitennavigation und Werbung auszusparen. Die so abgerufenen Daten von verschiedenen Quellen können sich gegenseitig ergänzen (verschiedene Aspekte, verschiedene Regionen) oder widersprechen (abweichende Werte für den gleichen Aspekt in der gleichen Region). Um das folgende Reasoning des Systems zu vereinfachen, werden diese Fälle vorher in einem Fusionsschritt in einheitliche Datenwerte für Regionen und Aspekte transformiert und mit daraus resultierender Unschärfe und Unsicherheitswerten annotiert.

ISBN 978-3-88579-286-4

Für die eigentliche Beantwortung der Anfrage werden die als potentiell relevant erachteten Informationen in der selben Knowledge Base (KB) instanziiert, in der auch die Anfrage des Benutzers sowie sein Kontext und sein Benutzerprofil als semantische Beschreibung vorliegen. In dieser KB sind außerdem die vollständige PDL sowie Zusammenhänge und Phänomene aus dem Bereich der Umwelt, Geographie und Gesundheit spezifiziert. Für die Erstellung der Ontologie wurden derzeit verfügbare Ontologien (z.B. SWEET[3]) sowie automatisch und manuell extrahierte Konzepte aus Texten der genannten Domänen verwendet. Durch verfügbare und nicht an die Domäne angepasste Reasoner werden in der KB weitere Informationen abgeleitet. Allein durch das Vorhandensein der Information, dass der Benutzer z.B. allergisch gegen Birkenpollen ist, der Information, dass die Birkenpollenkonzentration zum Zeitpunkt der Aktivität hoch ist, und der entsprechenden Zusammenhänge in der Ontologie, können relevante Folgen für den Benutzer abgeleitet werden.

Durch den anfangs bewusst weit gefassten Datenbestand und dem Reasoning enthält die resultierende Menge an Informationen auch viele triviale und irrelevante Inhalte. Für die Auswahl der relevantesten Inhalte wird ein unter Berücksichtigung von Anfrage, Kontext und Profil trainierter Klassifikator eingesetzt. Schließlich wird für diese Inhaltselemente der Modus der Präsentation gewählt (textuell oder graphisch) und eine Antwort generiert.

3.4 Ergebnisvisualisierung

Als Antwort des Systems stehen der Benutzungsoberfläche eine natürlichsprachliche Beschreibung der Situation sowie die für deren Erzeugung verwendeten Daten zur Verfügung. Um die Interpretation der Systemantwort für den Nutzer möglichst einfach zu gestalten sollte die Ergebnisanzeige verschiedene Anforderungen erfüllen. Das Ergebnis sollte im selben Kontext wie die Anfrage des Nutzers präsentiert werden. Außerdem soll die Anzeige von konkreten Werten möglichst verständlich, anpassbar und leicht zu überlagern sein. D.h. wenn mehrere Kenngrößen (z.B. Temperatur und Ozon) gemeinsam in einem Gebiet angezeigt werden sollen muss entschieden werden, welche Größen auf welche visuellen Strukturen abgebildet werden können, ohne die Interpretierbarkeit der anderen Anzeigen zu beeinträchtigen.

Alle Visualisierungen des Ergebnisses können sich an einer Managerkomponente als Kartenoverlay, Dialogfenster oder Steuerelement in der schon vorhandenen Oberfläche registrieren. Um die einfache Überlagerbarkeit zu gewährleisten wurden ein generisches Datenmodell und darauf abgestimmte Visualisierungstechniken entwickelt. Das Datenmodell kann numerische und qualitative Daten sowie Wertebereiche, Unsicherheit, und aus elementaren Werten zusammengesetzte Objekte abbilden. Die entsprechenden Visualisierungen sind meist für mehrere Teile dieses Datenmodells geeignet, so dass auch mehrere numerische Werte an einer Stelle gemeinsam durch alternative Visualisierungstechniken dargestellt werden können. Dies erlaubt, dass die Darstellungsweise für jeden Umweltaspekt vom System und den Nutzern für jede Sitzung individuell konfiguriert werden kann.

[3] http://sweet.jpl.nasa.gov/

ISBN 978-3-88579-286-4

Des Weiteren ist die Dimensionalität der Daten von Bedeutung bei der Auswahl der passenden Visualisierung. Konkrete Daten können an einem Punkt, entlang einer Route, in einer Fläche sowie über einen Zeitraum veränderlich vorliegen. Beispielhaft könnte der Ergebnistext durch ein mit Tabellen oder Symbolen angereichertes Dokument in einem Dialogfenster angezeigt werden, während Temperaturwerte in einem Bereich als Heatmap, und Ozonwerte entlang der Route als Balken dargestellt werden.

Stellt der Nutzer fest, dass das Ergebnis nicht seinen Erwartungen entspricht, hat er verschiedene Möglichkeiten in weiteren Anfrage-Iterationen Einfluss auf die Ergebniserzeugung zu nehmen. Zunächst kann er die gewählte Darstellung der verfügbaren Daten anpassen, falls eine aus seiner Sicht ungünstige Kombination aus Visualisierungen generiert wurde. Hierzu kann er sich nach Bedarf eine Matrix anzeigen lassen, welche die visuellen Mappings der Datensätze auf Visualisierungen (z.B. „Temperatur als Heatmap“, „Pollenbelastung als Balkendiagramm“) enthält. Die vom System erzeugte Auswahl lässt sich an dieser Stelle neu konfigurieren und andere Zuordnungen können ausgewählt werden. Dabei merkt sich das System die vorgenommenen Änderungen als Präferenzen des Nutzers und versucht sich bei der zukünftigen Ergebnisgenerierung daran zu orientieren.

Weiterhin soll der Nutzer bei Bedarf auf die Auswahl und Orchestrierung der verfügbaren Datenquellen Einfluss nehmen können. Eine Visualisierung basierend auf Graphen befindet sich momentan in der Entwicklung, mit der es ermöglicht werden soll, die Generierung des Resultats als Baum bestehend aus einzelnen Datenknoten darzustellen. Der Nutzer kann sich hierbei Details zu einem Datenknoten ausgeben lassen und einzelne Knoten nach Wunsch ausschließen, an anderer Stelle einbringen, durch automatisch vorgeschlagene Alternativen ersetzen sowie auch den Ablauf der Orchestrierung insgesamt verändern.

Zuletzt hat der Nutzer auch die Möglichkeit sein Profil und die eigentliche Anfrage zu modifizieren und erneut vom System auswerten zu lassen. Die verschiedenen Anfragewidgets sind dabei so ausgelegt, dass auf eine einfache und nachvollziehbare Weise Änderungen in die bestehende Anfrage eingebracht werden können. So stellt beispielsweise das Widget zur Routenselektion eine Auswahl der zuletzt selektierten Routen zur Verfügung, während sich ein Widget zur Auswahl des Verkehrsmittels die individuellen Präferenzen des Nutzers merkt und alternative Vorschläge, sortiert nach diesen Präferenzen, anbietet.

4 Fazit

PESCaDO ist eine neue Art von personalisierten Decision Support System, das mit Hilfe von semantischen Strukturen auf im Internet verfügbaren Daten arbeitet. Dabei wird mit Hilfe semantischer Technologien und der beschriebenen PDL das Problem gelöst, die Komplexität einer personalisierten Entscheidungsunterstützung, welche sich beliebige Datenquellen im Web in generischer Weise erschließt und dem Anwender in Form angepasster Visualisierungen zur Verfügung stellt, vor dem Nutzer zu verbergen und darauf aufbauend adaptive und effiziente Interaktionsmechanismen zur Verfügung zu stellen.

In diesem Beitrag wurde ein möglicher Aufbau eines solchen Systems und die notwendigen Schritte von Anfrageerstellung bis zur Interpretation der Antwort beschrieben. Durch

die anfrage- und kontextabhängige Auswahl der Datenquellen entsteht mehr als ein einfaches Daten-Mashup. Eine zentrale Beschreibung in Form einer Domänenontologie ist dabei ein wichtiger Bestandteil. Nachdem die Problembeschreibung und die Datenbasis in diese gemeinsame Basis transformiert wurden, können automatisiert eine Vielzahl an Aussagen generiert werden. Durch intelligente Klassifikation können sinnvolle Antworten durch Selektion, Textgenerierung und Visualisierung erstellt werden.

Die Antwort eines Systems ist natürlich immer nur so gut wie seine Datenbasis. Derzeit ist die Datenqualität aufgrund der Extraktion aus Webseiten noch nicht abzuschätzen. Doch auch bei sehr guten Ergebnissen wird die räumliche Auflösung der Daten gering sowie deren Verwendung in anderen Kontexten als der Webseite selbst aus urheberrechtlichen Gründen problematisch sein. Wir gehen jedoch von einer wachsenden Verfügbarkeit von strukturierten Daten im Internet aus und mit einer Steigenden Anzahl an von PESCaDO indizierten Diensten, wird die Qualität der Aussagen ebenfalls steigen.

Der PESCaDO Prototyp umfasst derzeit weite Teile der Benutzungsoberfläche zur Erstellung von Anfragen (siehe Abb. 2), die Serverkomponenten für deren Bearbeitung, sowie einfache Visualisierungen der Ergebnisse. Er ermöglicht damit einen kompletten Durchlauf des in Abb. 1 skizzierten Prozesses, wobei noch nicht alle Möglichkeiten der Personalisierung, insbesondere bei der Rückkopplung, berücksichtigt werden. Der Prototyp dient damit als Grundlage für die Forschung im Bereich der intelligenten Dialogsteuerung, kombinierter Visualisierung für komplexe Ergebnisse und der Orchestrierung von Diensten durch Benutzerinteraktion. Eine erste Demonstration vor potentiellen Nutzergruppen bestätigte den Bedarf an integrierter und visueller Datenrepräsentation sowie interaktiver Steuerung bei der personalisierten Präsentation von Umweltdaten, jedoch ist die Komplexität des Systems nicht unbedingt für klar umrissene Tätigkeiten, wie z.B. der täglichen Abfrage der aktuellen Pollenbelastung, gerechtfertigt. Außerdem hat sich ein starkes Interesse an mobilen Lösungen und regelmäßig zugesandten Berichten aus persönlich erstellten Anfragen gezeigt. Mit einer auf gewisse Aktivitäten und Regionen beschränkten Funktionalität, wird der Prototyp demnächst der Öffentlichkeit zur Verfügung gestellt[4]. Die weiteren Schritte bei der Entwicklung der Benutzungsoberfläche des PESCaDO Prototyps werden sich auf den Einbau von ad-hoc Visualisierungen während der Anfrageerstellung, sowie die intelligente Verknüpfung der Eingabemöglichkeiten konzentrieren.

Das PESCaDO Projekt wird gefördert durch die Europäische Kommission im siebten Rahmenprogramm (FP7-248594). Die Autoren möchten folgenden Personen für deren Mitarbeit bei der Entwicklung von PESCaDO danken: Leo Wanner, Nadjet Bouayad-Agha, Ulrich Bügel, Gerard Casamayor, Ari Karppinen, Ioannis Kompatsiaris, Tarja Koskentalo, Simon Mille, Jürgen Moßgraber, Anastasia Moumtzidou, Maria Myllynen, Emanuele Pianta, Marco Rospocher, Horacio Saggion, Luciano Serafini, Virpi Tarvainen, Sara Tonelli, Thomas Usländer, Stefanos Vrochidis.

[4] http://www.pescado-project.eu

Literatur

[CCLB97] Tiziana Catarci, Maria Francesca Costabile, Stefano Levialdi und Carlo Batini. Visual Query Systems for Databases: A Survey. *J. Visual Languages & Computing*, 8(2):215–260, 1997.

[DFT04] Paolo Dongilli, Enrico Franconi und Sergio Tessaris. Semantics Driven Support for Query Formulation. In *International Workshop on Description Logics*, Seite 104, 2004.

[DL01] C.-S.J. Dong und G.S.L. Loo. Flexible web-based decision support system generator (FWDSSG) utilising software agents. In *Proc. 12th Intl. Workshop on Database and Expert Systems Applications.*, Seiten 892 –897, 2001.

[OKI04] Satoshi Oyama, Takashi Kokubo und Toru Ishida. Domain-specific web search with keyword spices. *IEEE Transactions on knowledge and data engineering*, Seiten 17–27, 2004.

[Wöb06] Karl Wöber. Domain specific search engines. *Destination Recommendation Systems: Behavioral Foundations and Applications*, Seiten 205–226, 2006.

[WNBA+07] Leo Wanner, Daniel Nicklaß, Nadjet Bouayad-Agha, Bernd Bohnet, Joachim Bronder, Franciso Ferreira, Rainer Friedrich, Ari Karppinen, Francois Lareau, Achim Lohmeyer, Alejandra Panighi, Sergi Parisio, Helmut Scheu-Hachtel und Joao Serpa 12. From Measurement Data to Environmental Information - MARQUIS - A Multimodal AiR QUality Information Service for the General Public. In *Proceedings of Environmental Software Systems: Dimensions of Environmental Informatics*, Jgg. 7, Prague, 2007.

erschienen im Tagungsband der INFORMATIK 2011
Lecture Notes in Informatics, Band P192
ISBN 978-3-88579-286-4
weitere Artikel online:
http://informatik2011.de/519.html