Bundesministerium
für Wirtschaft
und Technologie

THESEUS
Forschungsprogramm für eine
neue internetbasierte Wissensinfrastruktur

Innovationspolitik, Informationsgesellschaft, Telekommunikation

Das THESEUS-Forschungsprogramm

Neue Technologien für das Internet der Dienste

www.bmwi.de
www.theseus-programm.de

Impressum

Redaktion & Gestaltung
Scholz & Friends Agenda

Druckerei
Richard Thierbach, Buch- und Offset-Druckerei

Bildnachweis
BMWi, Corbis, Getty, Fotolia, Masterfile

Herausgeber
Bundesministerium für Wirtschaft und Technologie
Kommunikation und Internet (P3)
10115 Berlin

In Zusammenarbeit mit:

THESEUS Programmbüro
c/o empolis GmbH
An der Autobahn
Postfach 180
33311 Gütersloh

Stand
November 2008

Zertifikat seit 2002
audit beruf**und**familie

Das Bundesministerium für Wirtschaft und Technologie ist mit dem audit berufundfamilie® für seine familienfreundliche Personalpolitik ausgezeichnet worden. Das Zertifikat wird von der berufundfamilie eGmbH, einer Initiative der Gemeinnützigen Hertie-Stiftung, verliehen.

Bundesministerium
für Wirtschaft
und Technologie

THESEUS
Forschungsprogramm für eine
neue internetbasierte Wissensinfrastruktur

Innovationspolitik, Informationsgesellschaft, Telekommunikation

Das THESEUS-Forschungsprogramm

Neue Technologien für das Internet der Dienste

Inhalt

Vorwort 05

1 THESEUS – neue Technologien für das Internet der Dienste 06

2 ALEXANDRIA – Wissensplattform im Internet 08

3 CONTENTUS – Technologien für die Mediathek der Zukunft 10

4 MEDICO – intelligente Recherche in Medizindatenbanken 12

5 ORDO – Ordnung für die digitale Welt 14

6 PROCESSUS – Wissen in Unternehmen besser nutzen 16

7 TEXO – Infrastruktur für internetbasierte Dienste 18

8 THESEUS-Basistechnologien 20

9 THESEUS-Begleitforschung 23

10 THESEUS Talente 2008 – Ideenwettbewerb zur Nachwuchsförderung 24

11 THESEUS Mittelstand 2009 – Technologiewettbewerb für den Mittelstand 25

12 Die Partner des THESEUS-Programms 26

Vorwort

IKT – hinter diesen drei Buchstaben verbirgt sich die Zukunft. Mit den Informations- und Kommunikationstechnologien (IKT) können wir auch morgen Wohlstand und Arbeitsplätze sichern. Deshalb fördern wir die Entwicklung genau auf diesem Gebiet. Mein Haus, das Bundesministerium für Wirtschaft und Technologie (BMWi), hat deshalb das THESEUS-Forschungsprogramm aufgelegt. Es verknüpft in herausragender Weise wissenschaftliche Exzellenz und wirtschaftliches Potenzial. THESEUS zielt auf neue Technologien für das Internet der Dienste. Auf dem ersten IT-Gipfel der Bundesregierung wurde das Programm zu einem Leuchtturmprojekt erklärt.

Nun werden auf dem dritten IT-Gipfel erste Ergebnisse von THESEUS vorgestellt. Im THESEUS-Programm sind heute Marktführer aus den Bereichen Unternehmenssoftware, fertigende Industrie und Medizintechnik beteiligt. Ein Punkt ist mir besonders wichtig: Im Programm spielen nicht nur die Großen eine Rolle, THESEUS richtet sich gerade auch an kleine und mittlere Unternehmen. Sie werden mit der ergänzenden Fördermaßnahme THESEUS-MITTELSTAND angesprochen, sich mit Projektideen zu beteiligen.

Mehr noch: Im Rahmen von THESEUS arbeiten führende Forschungseinrichtungen und Elite-Universitäten an der Entwicklung semantischer Technologien, um komplexes Wissen im Internet besser erschließen und nutzen zu können. Aber nicht nur die Spitzenforschung ist uns wichtig, auch der Nachwuchs liegt uns am Herzen. Mit dem Ideenwettbewerb THESEUS TALENTE öffnete sich das Forschungsprogramm für junge Talente, die am Internet der Zukunft mitarbeiten wollen.

THESEUS ist mit einem Gesamtvolumen von ca. 200 Mio. Euro, von dem insgesamt ca. 100 Mio. Euro vom BMWi beigesteuert werden, das größte IKT-Forschungsprojekt der Bundesregierung. Mit THESEUS wollen wir eine Spitzenposition in den Informations- und Kommunikationstechnologien erlangen. Dies nutzt dem Standort Deutschland insgesamt.

Für den Bürger werden die THESEUS-Ergebnisse von großem Nutzen sein, da ein besserer Wissenszugang ermöglicht wird und neue, effizientere Dienste im Internet bereitgestellt werden. Die Wirtschaft nutzt die THESEUS-Ergebnisse, um ihre Geschäftsprozesse zu optimieren und neue Dienste und Geschäftsmodelle zu entwickeln. Damit werden neue Märkte entstehen und neue Geschäftsfelder können erschlossen werden. Letztendlich wollen wir mit unserer Förderung Anreize für den Ausbau von Forschung und Entwicklung in den Unternehmen schaffen, um das Wirtschaftswachstum und die Schaffung von Arbeitsplätzen anzuregen. Schließlich werden durch die Erforschung und weitere Entwicklung der Technologien für das Internet der Dienste Impulse für den Forschungssektor gegeben.

Die vorliegende Broschüre gibt Ihnen einen Überblick über das THESEUS-Programm. Beteiligen Sie sich an dem Programm! Sie können davon profitieren.

Ihr

Michael Glos
Bundesminister für Wirtschaft und Technologie

1
THESEUS – neue Technologien für das Internet der Dienste

Das Internet bietet täglich neue Informationen. Mit der zunehmenden Datenmenge stellt sich aber auch die Herausforderung, Nutzern einen einfachen und effizienten Zugang zu diesem immensen Wissensschatz zu ermöglichen. Im Rahmen von THESEUS, dem derzeit größten deutschen IT-Forschungsprogramm, arbeiten Partner aus Wissenschaft und Wirtschaft gemeinsam an der Lösung dieser Aufgabe.

Wie die griechische Sagengestalt Theseus einst den Weg aus dem Labyrinth des Minotaurus fand, so sollen die im gleichnamigen Forschungsprogramm entwickelten Ansätze den Weg durch die wachsende Datenmenge des Internets weisen. Unter dem Dach von THESEUS entwickeln 30 Forschungspartner aus Wissenschaft und Wirtschaft neue Technologien und Anwendungen. Das Ziel: THESEUS soll den Zugang zu Informationen vereinfachen, Daten zu neuem Wissen vernetzen und die Grundlage für die Entwicklung neuer Dienstleistungen im Internet schaffen.

Die Technologien, die im Rahmen von THESEUS entwickelt werden, haben großes Potenzial, denn sie können das sogenannte Internet der Dienste ermöglichen. Während heute Dienstleistungen wie Online-Shopping, Flugbuchungen oder Recherchedienste nur einzeln im Netz verfügbar sind, können im Internet der Dienste solche Angebote automatisch gebündelt und vernetzt werden. Künftig könnte ein Nutzer dem Computer einfach mitteilen: „Ich möchte von Berlin nach Hamburg ziehen.“ Das Computerprogramm würde eigenständig die passenden Angebote für die Wohnungssuche, den Umzug bis hin zum Anmelden des Wohnsitzes ermitteln und koordinieren.

Semantische Technologien als Grundlage

Damit diese Vision Wirklichkeit wird, werden im THESEUS-Programm neue semantische Technologien entwickelt. Mit diesen lassen sich Informationen auf ihre Inhalte hin auswerten, einordnen und verknüpfen. Computerprogramme können damit Informationen nicht nur, wie heute üblich, mit Hilfe von Schlagwörtern oder Inhaltsfragmenten finden, sondern auch eigenständig deren Bedeutung ermitteln, sie in Beziehung zu anderen Informationen setzen, als Ordnungssysteme modellieren – sogenannte Ontologien – und nach bestimmten Regeln logische Schlüsse daraus ziehen.

Eine wichtige Grundlage für diesen Fortschritt ist die Entwicklung von Basistechnologien. Im Rahmen von THESEUS arbeiten die beteiligten Forscher zum Beispiel an innovativen Verfahren zur automatischen Erzeugung von Metadaten – vergleichbar mit Schlüsseldaten oder Indizes – zur schnelleren Verarbeitung von multimedialen Dokumenten und an neuen grafischen Benutzeroberflächen.

Erprobung in Anwendungsszenarien

Damit auf Grundlage der Basistechnologien möglichst schnell neue Produkte und Dienstleistungen

entstehen, werden sie von den beteiligten Partnern in sechs Anwendungsszenarien prototypisch eingesetzt und erprobt. In diesen Szenarien wird ermittelt, wie die Technologien für innovative Werkzeuge, Dienste und Geschäftsmodelle genutzt werden können.

Im Anwendungsszenario ALEXANDRIA wird der sogenannte Web-2.0-Ansatz – die Vernetzung der Nutzer in virtuellen Gemeinschaften und die wachsende Bedeutung von nutzergenerierten Inhalten – um innovative Methoden des Wissensmanagements erweitert. Damit sollen Nutzer einfacher und individueller auf die Inhalte im Internet zugreifen können.

Das Anwendungsszenario CONTENTUS nutzt semantische Technologien für die Digitalisierung von Kulturgütern in Form von Texten, Bildern sowie Ton- und Videoaufzeichnungen. Ziel ist es, Kulturgüter für die Nachwelt besser zu bewahren und medienübergreifend einen schnellen und umfassenden Zugang zu diesem Wissen zu ermöglichen.

Im Rahmen des Anwendungsszenarios PROCESSUS wird daran geforscht, wie das in Unternehmen verfügbare Wissen besser genutzt werden kann. Dazu wird eine Plattform entwickelt, die Ressourcenplanung und das Management von Geschäftsprozessen integriert und den Vergleich von Produkten, Lösungen und Geschäftspartnern erleichtert. So können Unternehmen ihre Kosten senken und den gesamten betrieblichen Ablauf optimieren.

Im Anwendungsszenario MEDICO soll die Qualität der medizinischen Diagnostik auf Basis bildgebender Verfahren verbessert werden. Semantische Technologien unterstützen den Arzt dabei, schnell und sicher Auffälligkeiten in medizinischem Bildmaterial zu erkennen und seine Diagnose und Therapieentscheidungen durch eine schnelle Auswertung von Bilddatenbanken und Fachliteratur zu verbessern.

Das Anwendungsszenario ORDO konzentriert sich auf die Entwicklung von innovativen Ansätzen und Technologien für die Ordnung von digitalen Informationen. Ziel der Forschung sind Werkzeuge, die Informationen automatisch ordnen und so die Handhabung und Auswertung von großen Datenmengen erleichtern. Davon können unter anderem jene profitieren, die tagtäglich mit großen, schwer überschaubaren Datenbeständen, wie zum Beispiel Patentinformationen, arbeiten.

Damit die Angebote im künftigen Internet der Dienste in einem offenen und dennoch sicheren Umfeld genutzt werden können, entwickeln die Forscher im Anwendungsszenario TEXO eine integrierte Plattform für die Bereitstellung, Abwicklung und Kombination von internetbasierten Dienstleistungen. Sie wird es sowohl Anbietern als auch Nutzern erleichtern, diese Dienste zu entwickeln und in Anspruch zu nehmen, und ist ein wichtiger Meilenstein auf dem Weg zum Internet der nächsten Generation.

Impulse für den Standort Deutschland

Mit der Forschung im THESEUS-Programm werden neue Technologien für das Internet der Dienste geschaffen und erstmals auch komplexes Wissen abgebildet. Die beteiligten Forscher ermöglichen so die einfache Nutzung des im Internet verfügbaren Wissens und schaffen die Grundlage für die Entwicklung von neuen Anwendungen. THESEUS stärkt auf diese Weise nicht nur Deutschlands Position im Bereich der Internetforschung, sondern trägt auch dazu bei, dass Unternehmen aus Deutschland das Internet der Dienste mit innovativen Produkten und Dienstleistungen aktiv mitgestalten und von diesem Zukunftsmarkt profitieren.

Im Rahmen von Wettbewerben lädt THESEUS externe Wissenschaftler und mittelständische Unternehmen ein, sich in einer frühen Phase an diesem Programm zu beteiligen und an den ersten Ergebnissen zu partizipieren.

THESEUS hat eine Laufzeit von fünf Jahren und wird vom Bundesministerium für Wirtschaft und Technologie (BMWi) mit rund 100 Mio. Euro gefördert. Zusätzliche 100 Mio. Euro werden von den beteiligten Partnern aus Industrie und Forschung aufgebracht, sodass insgesamt rund 200 Mio. Euro in die zukunftweisenden Forschungsarbeiten fließen – Technologien für das Internet der Dienste und den Standort Deutschland.

2
ALEXANDRIA – Wissensplattform im Internet

In der gleichnamigen Stadt an der ägyptischen Mittelmeerküste lag die berühmteste Bibliothek der Antike. Ihr präzises System der Katalogisierung vereinfachte schon damals den Zugang zu Wissen. Inspiriert durch dieses Vorbild entwickeln die Forscher im Rahmen von ALEXANDRIA eine Wissensplattform im Internet, die durch völlig neue Interaktionsmöglichkeiten die Erschließung von Wissen erleichtert.

Ausgangspunkt des Anwendungsszenarios ALEXANDRIA ist die wachsende Bedeutung der Internetnutzer im sogenannten Web 2.0. Heute nutzen Anwender nicht nur Suchmaschinen, E-Mail- und Chat-Dienste, sondern stellen im Internet selber aktiv ihr Wissen zur Verfügung. Bekannte Beispiele für diese neue Form der Interaktion sind Blogs, die Internetenzyklopädie Wikipedia oder soziale Netzwerke wie Facebook oder Xing. Diese Plattformen wären ohne Nutzer, die Wissen und Informationen bereitstellen und austauschen, nicht denkbar.

Wissen sammeln und zugänglich machen

Ziel der an ALEXANDRIA beteiligten Forscher ist es, das im Rahmen des Web 2.0 verfügbare Wissen besser zu vernetzen und leichter zugänglich zu machen. Dazu wird eine Wissensplattform entwickelt, die nicht nur Informationen sammelt und dabei das von Nutzern generierte Wissen einbezieht, sondern auch dazu beiträgt, die tägliche wachsende Datenmenge besser zu verwalten. Diese Wissensplattform soll die Nutzer bei der Veröffentlichung und Weiterentwicklung von Inhalten sowie bei der Suche nach Inhalten unterstützen. Die Platt-

form soll einfach und intuitiv bedienbar gestaltet werden. Dazu werden semantische Technologien und Ordnungssysteme entwickelt, mit denen auch große Datenmengen verarbeitet und sinnvoll miteinander verknüpft werden können.

Das Wissen der Nutzer wird einbezogen

Weiteres Ziel von ALEXANDRIA ist es, die Internetnutzer und ihr Wissen als integralen Bestandteil der geplanten intelligenten Wissensplattform einzubeziehen. Ein Beispiel dafür, wie dies erfolgen kann, ist das sogenannte Tagging. Damit wird die Zuordnung von Schlagwörtern zu Inhalten, etwa Dokumenten oder Bildern, bezeichnet. Eine solche Anreicherung von Daten – auch als Metadaten bezeichnet – kann sowohl von Menschen durchgeführt werden als auch automatisch erfolgen. So können technologische Lösungen zum Beispiel Schreibfehler automatisch erkennen und verbessern, Menschen hingegen können schneller und besser Fehldeutungen ähnlicher oder gleicher Begriffe korrigieren. Die im Anwendungsszenario ALEXANDRIA entstehende Wissensplattform soll die spezifischen Stärken beider Ansätze kombinieren.

Umfassender Zugriff auf Wissen

Um die Möglichkeiten einer solchen Wissensplattform zu demonstrieren, widmet sich ALEXANDRIA in einer ersten Beispielanwendung dem Thema „Berühmte deutsche Persönlichkeiten“. Dabei wird auf einer einfach zu navigierenden Benutzeroberfläche demonstriert, wie der Zugang zu Wissen mit den entwickelten Technologien erleichtert wird. Auf einem Bildschirm werden ausgewählte Persönlichkeiten präsentiert. Möchte sich der Nutzer schnell informieren, wer eine der dargestellten Personen war, so kann er automatisch extrahierte Informationen und fünf wichtige Fakten lesen. Welche fünf Fakten als wichtig präsentiert werden, entscheiden die Nutzer in einem kontinuierlichen Abstimmungsprozess, in dem sie die Relevanz der Daten individuell bewerten. Auch weitergehende Informationen, die dem Nutzer angeboten werden, können detailliert bewertet werden. Durch diesen Einbezug der Nutzer und ihres Wissens wird die Informationsqualität der präsentierten Inhalte erhöht und sichergestellt, dass tatsächlich die aus Nutzersicht wichtigsten Daten angezeigt werden.

Um die Möglichkeiten der inhaltlichen Verknüpfung von Daten zu demonstrieren, zeigt außerdem ein „Beziehungsbrowser“ Beziehungen der dargestellten Persönlichkeiten zu anderen berühmten Zeitgenossen an. Die präsentierten Beziehungen werden automatisch aus den Datenbeständen extrahiert. Die Nutzer haben die Möglichkeit, Beziehungen, die durch die automatischen Verfahren fehlerhaft oder gar nicht erkannt wurden, zu berichtigen und zu ergänzen.

Dieses Beispiel zeigt die wichtige Rolle, die dem Nutzer bei der Erstellung, Anreicherung und Bewertung von Inhalten im Rahmen von ALEXANDRIA zukommt. Mit den entwickelten Technologien soll die Erschließung dieses Nutzerwissens durch semiautomatische Verfahren vereinfacht und so dazu beigetragen werden, Informationen so zu ordnen, wie sie benötigt werden: das Wichtigste zuerst.

Ziele von ALEXANDRIA

- Einfache und intuitiv bedienbare Wissensplattform im Internet entwickeln
- Nutzer aktiv einbinden
- Nutzer bei der Veröffentlichung und Weiterentwicklung von Inhalten unterstützen
- Kombination von manueller und automatischer Metadatenerzeugung

3
CONTENTUS – Technologien für die Mediathek der Zukunft

Millionen von Büchern, Bildern, Tonbändern und Filmen lagern in Deutschlands rund 30.000 Bibliotheken, Museen und Archiven – ein unermesslicher Wissensschatz. Den Fragen, wie dieses kulturelle Erbe einer möglichst großen Zahl von Menschen zugänglich gemacht und wie es für nachfolgende Generationen bewahrt werden kann, gehen die an CONTENTUS beteiligten Forscher nach.

Es gilt als eine der großen Aufgaben der Wissensgesellschaft, Nutzern einen einfachen und medienübergreifenden Zugriff auf Wissens- und Kulturgüter zu ermöglichen. Im Rahmen von CONTENTUS werden Konzepte und Technologien für eine Infrastruktur entwickelt, die es Kultureinrichtungen und Informationsanbietern erlaubt, diesem Ziel näherzukommen. Diese neuen Technologien ermöglichen die automatische Aufbereitung und semantische Vernetzung großer Datenmengen, sei es in der Form von Texten, Bildern oder Video- und Tonaufnahmen. Die Arbeiten von CONTENTUS sind eng abgestimmt und verknüpft mit der Initiative „Deutsche Digitale Bibliothek“ der Bundesregierung. Hierbei ist eine weitreichende Beschreibung der Inhalte durch sogenannte Metadaten wichtig – vergleichbar mit Schlüsseldaten oder Indizes zum leichten Auffinden. Die Vernetzung der verschiedenen digitalen Wissensbestände im Internet und die Einbindung von nutzergenerierten Inhalten stellt eine weitere Herausforderung dar. Hierfür entwickeln die Forscher intelligente Algorithmen, die alle notwendigen Prozesse automatisieren und den Nutzern die Möglichkeit eröffnen, digitale Inhalte mit eigenen Informationen anzureichern. Das Ergebnis: untereinander vernetzte virtuelle Mediatheken der nächsten Generation.

Next Generation Multimedia Libraries

Die Forscher entwickeln im Rahmen von CONTENTUS Konzepte und Technologien, auf deren Grundlage sogenannte Next Generation Multimedia Libraries realisiert werden können. Diese „Multimediatheken“ vernetzen die Sammlungen von Bibliotheken, Medienarchiven und Sendeanstalten zu einer neuen Informationsstruktur, die Anbieter und Nutzer über das Internet zusammenführt. Dabei werden neue Möglichkeiten geschaffen, damit die bestehenden multimedialen Bestände um das Wissen der Nutzer ergänzt werden können. Zugleich werden die Daten semantisch, also nach ihrem inhaltlichen Zusammenhang, zu einem neuartigen Wissensnetz verbunden. Auf diese Weise können Anbieter ihre multimedialen Sammlungen in optimaler Struktur und Qualität bereitstellen.

Ein Beispiel verdeutlicht den Nutzen dieser Multimediathek der nächsten Generation: Eine Recherche zum Thema „Lukullus“ liefert zahlreiche medienübergreifende Suchergebnisse, wie Texte, Bilder, Filme oder Hörbeispiele. Dank der Verwen-

dung semantischer Technologien erscheinen auch Inhalte, die in einem Sinnzusammenhang zu dem Thema stehen. Daher können Partituren oder Aufnahmen zur „Verurteilung des Lukullus“ ebenso angezeigt werden wie andere Stücke des Autors Bertolt Brecht und des Komponisten Paul Dessau, beispielsweise „Der gute Mensch von Sezuan“. Es werden auf Wunsch auch Informationen und Musiktitel der großen Brechtinterpreten angezeigt oder Verweise auf kommerzielle Musikanbieter, bei denen man Brecht’sche Lieder online kaufen kann. Während der Nutzer klassischer Archive meist erst nach langer Suche auf einzelne Werke oder Sammlungsstücke stößt, profitiert der Nutzer der geplanten Multimediathek von einer schnellen Suche und von sinnvollen, medienübergreifenden Ergebnissen.

Vorteile für Nutzer und Anbieter

Mit CONTENTUS können Nutzer künftig einfacher im digitalisierten Kulturerbe recherchieren und navigieren. Sie können dabei auch auf einfache Weise Ergebnisse und Erkenntnisse mit anderen Nutzern und Anbietern teilen und Inhalte selbstständig durch Verknüpfungen anreichern. Im Rahmen von CONTENTUS wird auch eine spezielle Plattform für wissenschaftliches Arbeiten mit digitalen Inhalten geschaffen, mit der Anbieter, etwa kulturelle Einrichtungen, ihre multimedialen Sammlungen einem breiten Nutzerkreis zugänglich machen können. Dazu werden Techniken für die qualitative Aufbereitung und semantische Vernetzung von multimedialen Beständen geschaffen, die es auch kleineren Bibliotheken und Archiven erlauben, Teil der geplanten Informationsstruktur zu werden.

Forschung und Entwicklung im Dienst des Wissens

Durch CONTENTUS kommen zahlreiche Technologien und neue Konzepte erstmals zum Einsatz, die zum Aufbau eines großen Wissensnetzes nötig sind. In sechs Arbeitsschritten werden dabei analoge Datenträger für die Next Generation Multimedia Libraries digital aufbereitet. Zunächst helfen sogenannte Hochdurchsatzverfahren dabei, große Mengen analoger Inhalte schnell und effizient zu digitalisieren. Im zweiten Schritt entwickeln die Forscher Technologien, um die Qualität des eingelesenen Materials zu messen und zu optimieren. Im Anschluss daran werden auf Basis der Inhalte durch Verfahren der Text-, Stichwort-, Stimm- oder Spracherkennung Metadaten zur Beschreibung der Inhalte erzeugt – ohne diese Art von Indizierung wären digitalisierte Bücher, Filme oder Tonaufnahmen in den Weiten des Internets nicht mehr lokalisierbar. Auf der Grundlage dieser Metadaten lassen sich Inhalte im Rahmen der semantischen Vernetzung in einem vierten Schritt miteinander verknüpfen. Wie in dem Recherchebeispiel zum „Lukullus“ werden dem Nutzer über einzelne gefundene Inhalte weitere sinnverwandte Informationen vorgeschlagen – und führen ihn, wenn gewünscht, auf eine thematische Erkundungsreise. Neben einer klassischen Volltextsuche werden so auch Recherchen ermöglicht wie „Welche der Kinder von Thomas Mann waren ebenfalls Schriftsteller und welches sind ihre Werke?“. Der fünfte Schritt bezieht Benutzer und Experten mit ein, die Inhalte ergänzen können. Der abschließende Schritt komplettiert den Prozessablauf: der vernetzte und medienübergreifende Zugriff auf Informationen – ein Meilenstein in der Weiterentwicklung der internetbasierten Wissensinfrastruktur.

Ziele der CONTENTUS-Anwendungen

- Kulturelles Erbe sichern
- Konzepte und Technologien für sogenannte Next Generation Multimedia Libraries entwickeln
- Nutzergenerierte Inhalte einbinden
- Neue semantische und medienübergreifende Recherchemöglichkeiten schaffen

4
MEDICO – intelligente Recherche in Medizindatenbanken

Bildgebende Verfahren, wie zum Beispiel Ultraschall oder Computertomographie (CT), sind eine wichtige Grundlage für die medizinische Diagnose und Therapie, denn sie ermöglichen es, Krankheiten frühzeitig zu erkennen und gezielt zu behandeln. Bislang gibt es allerdings kein Instrument, das all die damit verbundenen heterogenen Informationen – Texte, Bilder, Labordaten – intelligent strukturiert und zugänglich macht. Im Anwendungsszenario MEDICO arbeiten Forscher daran, diese Lücke zu schließen.

Wie kann ein Computer lernen, Bilder zu interpretieren, Ähnlichkeiten zu erkennen und sie mit Textinhalten in einen sinnvollen Zusammenhang zu setzen? Das sind die Fragen, mit denen sich das THESEUS-Anwendungsszenario MEDICO beschäftigt. Auf Basis semantischer Technologien entwickeln die beteiligten Forscher Anwendungen für eine einfache und vernetzte Recherche in Medizindatenbanken. Sie soll Ärzten und anderen im Gesundheitssektor tätigen Personen künftig die Arbeit erheblich erleichtern, indem sie alle relevanten Patienteninformationen aus bild- und textbasierten Befunden intelligent zusammenführt.

Neue Möglichkeiten für Diagnose und Therapie

MEDICO erkennt nicht nur anatomische Strukturen wie Knochen, Gefäße oder Organe – sondern katalogisiert auch Daten automatisch, trägt Vergleichsbilder und Behandlungsberichte aus mehreren Datenbanken zusammen und bemerkt so krankhafte Veränderungen. Am Beispiel der Diagnose und Therapie von Lymphomen lässt sich der Nutzen von MEDICO veranschaulichen: Ein Lymphompatient befindet sich in einem Kreiskrankenhaus in Behandlung. Der behandelnde Arzt will wissen, ob die Chemotherapie angeschlagen hat. MEDICO vergleicht automatisch aktuelle CT-Bilder mit solchen, die vor Therapiebeginn gemacht wurden, und extrahiert daraus relevante Informationen. Das Ergebnis: Die Chemotherapie scheint zu wirken, die Lymphknoten haben sich nicht weiter vergrößert. Aber dem Programm fallen neue Schädigungen in der Milz auf, die Fragen aufwerfen. Vor Ort hat der behandelnde Arzt keine vergleichbaren Fälle, die Hinweise liefern könnten. Zur Klärung macht er sich eine weitere Fähigkeit von MEDICO zunutze. Digital werden Bilder und Befunde seines Patienten an das Rechenzentrum einer großen Uniklinik geschickt. Das Programm gleicht die Daten des Lymphompatienten aus dem Kreiskrankenhaus

mit den zahlreichen Lymphomfällen ab, die in der Uniklinik behandelt wurden. Auf dieser um ein Vielfaches erweiterten Informationsgrundlage erhält der behandelnde Onkologe präzise Empfehlungen zu weiterführenden Untersuchungen und Therapien. Durch das Zusammenspiel mit der THESEUS-Komponente ORDO kann der Arzt darüber hinaus auch schnell eine Recherche in aktueller wissenschaftlicher Literatur durchführen, die sich mit Milzläsionen bei Lymphompatienten befasst. Durch die Einbeziehung aktuellster Forschungsergebnisse kann so die optimale Behandlung bestimmt werden.

Unterstützung für Ärzte und Patienten

Die Möglichkeiten, die sich aus der intelligenten Interpretation und Zusammenführung von unterschiedlichsten Bildern und Texten für den Medizinsektor ergeben, sind vielfältig. Der behandelnde Arzt bekommt eine effiziente und präzise Entscheidungshilfe an die Hand. Zugleich profitieren auch die Patienten davon, dass der behandelnde Arzt seine diagnostischen und therapeutischen Entscheidungen auf Basis eines großen Erfahrungsfundaments stellen kann. Zudem wird MEDICO Krankenhäusern ermöglichen, Zeit und Kosten zu sparen. Denn für den Patienten unnötige und oft beschwerliche Untersuchungen können vermieden werden.

Im Rahmen von MEDICO konzentrieren sich die beteiligten Forscher zunächst auf die semantische Suche und die Strukturierung von Informationen in medizinischen Bilddatenbanken. Das Anwendungspotenzial von MEDICO geht aber deutlich darüber hinaus. Wenn die ethischen und datenschutzrechtlichen Randbedingungen geklärt sind, können die Patienten sowie die pharmazeutische Industrie von einem schnellen und effizienten Zugriff auf relevante Studien profitieren, was die Qualität von Diagnose und Therapie weiter verbessern kann. Auch Versicherungen und Behörden können mit den entwickelten Technologien künftig wesentlich präzisere demografische oder statistische Untersuchungen auf der Grundlage anonymisierter Daten durchführen.

Daten sinnvoll strukturieren

Die mit MEDICO entwickelten Technologien sollen den Computer dazu befähigen, selbstständig Informationen aus dem Inhalt der dargestellten Bilder zu erzeugen, diese mit weiteren medizinischen Daten zusammenzuführen und zu strukturieren, um daraus für den behandelnden Arzt Empfehlungen abzuleiten. Um dieses Ziel zu erreichen, müssen noch zahlreiche Komponenten entwickelt werden, zum Beispiel Verfahren der Mustererkennung, computergestützte Erkennungssysteme oder klinische Entscheidungshilfen. Die zentrale Aufgabe besteht jedoch in der sogenannten Modellierung einer passenden Ontologie. Damit wird die Entwicklung eines Ordnungssystems bezeichnet, das die Fülle von Informationen auf Grundlage des medizinischen Fachvokabulars gliedern und vernetzen kann. Die dafür notwendigen Metadaten werden von MEDICO automatisch generiert und den jeweiligen Bilddaten zugeordnet. Bei MEDICO basiert die Strukturierung der Bildinhalte auf der menschlichen Anatomie. Zur Beschreibung der Bildinhalte werden von Medizinern entwickelte Ontologien wie „RadLex“ oder das sehr detaillierte „Foundational Model of Anatomy Ontology“ verwendet.

Durch die Entwicklung immer besserer Diagnose- und Behandlungsmethoden wächst die Informationsmenge, die dem Arzt über seine Patienten und deren Krankheitsbilder zur Verfügung steht. MEDICO interpretiert, verknüpft und vergleicht medizinische Bild- und Textinformationen schnell und einfach – und leistet so einen wichtigen Beitrag zum medizinischen Fortschritt.

Ziele von MEDICO

- Medizinische Informationen wie Bilder, Befunde und Texte intelligent recherchieren und strukturieren und Informationen für unterschiedliche Gesundheitssituationen bereitstellen
- Hilfe bei diagnostischen und therapeutischen Entscheidungen leisten
- Schneller und effizienter auf medizinisches Wissen im Klinikalltag und in der Forschung zugreifen

5
ORDO – Ordnung für die digitale Welt

Das Wissen der Menschheit liegt in vielen Formen vor. Informationen werden als Texte, Bilder, Video- oder Tonaufnahmen auf verschiedenen Datenträgern in unterschiedlichen Dateiformaten gespeichert. Im Rahmen des Anwendungsszenarios ORDO entwickeln Forscher ein System, das diese Vielfalt besser beherrschbar machen soll.

Die moderne Informationstechnologie macht das Leben der Menschen einfacher, aber manchmal auch unübersichtlicher. Wer zum Beispiel den Familienurlaub dokumentieren möchte, steht vor einem Problem: Die Fotos sind in der Digitalkamera gespeichert, die Strandgrüße stecken als SMS im Handy und die Tochter hat ihr Reisetagebuch in den Laptop getippt. Wie soll der Vater daraus eine sehenswerte Präsentation zaubern, die man Freunden zeigen kann? Wie soll er die Informationen übersichtlich ordnen und auswählen, was wichtig ist?

Wissen besser ordnen

Im Rahmen von ORDO sollen neue Dienste und Softwarewerkzeuge entstehen, die dem Nutzer eine solche Ordnung und Priorisierung von Informationen und Suchergebnissen erleichtert. Hierzu werden semantische Verfahren verwendet und eingesetzt. So werden Technologien entwickelt, die dem Nutzer alle Daten zur Verfügung stellen, die inhaltlich zu seiner Suchanfrage passen – und es ihm erlauben, Wichtiges von Unwichtigem zu unterscheiden, Zusammenhänge zu erkennen und Hintergrundinformationen einzusehen. Die neuen Technologien würden es dem Familienvater erlauben, auf einen Blick zu sehen, auf welchen Datenträgern sich welche Informationen über den Urlaub verbergen. Er könnte sie einfach einsehen, ihre Relevanz bestimmen und sich vielleicht entschließen, auf das Tagebuch der Tochter in der Präsentation doch zu verzichten.

Dieser Überblick über die verfügbaren Informationen wird automatisch generiert, denn die ORDO-Technologien räumen von selbst auf. Darüber hinaus werden Ansätze entwickelt, um große Datenmengen aufbereiten und grafisch darstellen zu können. So entsteht ein umfassendes und individuelles Wissensmanagement, das es dem Nutzer ermöglicht, stets die richtigen Informationen zu finden.

Praxistest in der Forschung und Entwicklung

Die Forschungsergebnisse von ORDO werden speziell für die Bedürfnisse von Forschungs- und Entwicklungsabteilungen, z. B. in der Chemie- und Pharma-Branche, und die dort notwendige Verarbeitung von großen Datenbeständen im Rahmen von Demonstratoren erprobt. Neben den primären Daten wie z. B. Patenten müssen in diesem Bereich auch umfangreiche Sekundärinformationen wie Forschungsberichte und andere wissenschaftliche Veröffentlichungen ausgewertet werden. Dazu kommen E-Mails, Archive, firmeninterne WiKis und Blogs. Forschungsziel von ORDO ist es, Methoden und Technologien zu entwickeln, die alle diese Datenbestände effizienter nutzbar machen und ihre Verknüpfung mit Inhalten aus dem Internet erlauben.

Umfassende Recherchemöglichkeiten

Ein Beispiel aus der Wissenschaft verdeutlicht den Nutzen der ORDO-Technologien: Ein Arzneimittelforscher sucht Informationen zu dem Wirkstoff Acetylsalicylsäure. Ein umfassendes Recherche-ergebnis sollte die Darstellung des Wirkstoffs als chemische Formel, seine Wirkungsweise und eine Auflistung von Arzneimitteln inklusive der Bilder konkreter Produkte, in denen er verwendet wird, umfassen. All diese Informationen muss der Forscher sich bisher einzeln zusammensuchen – aus einer chemischen Datenbank, aus einem Arzneimittelarchiv oder sogar aus Marketingunterlagen der Hersteller.

Mit Hilfe von ORDO könnte er seine Anfrage an eine zentrale Plattform richten, von der aus die Recherche koordiniert wird. Das eingesetzte Computerprogramm kann selbstständig die chemische Formel von Acetylsalicylsäure in einer Datenbank ermitteln und dann mit dieser Formel die passende Moleküldarstellung finden. Das Programm schließt aus dem Bild einer Aspirinpackung, auf der der Wirkstoff erwähnt wird, dass dieses Arzneimittel in diesem Zusammenhang wichtig ist – und erkennt auch, wofür der Wirkstoff eingesetzt wird. All diese Informationen werden dem Wissenschaftler in einem zweiten Schritt in einer übersichtlichen Darstellung präsentiert. Dabei kann der Forscher seine Rechercheergebnisse nach seinen Bedürfnissen ordnen und individuell auswählen, was für ihn relevant ist und worüber er noch mehr erfahren möchte.

Informationsmanagement durch semantische Technologien

Mit ORDO werden Computer bald auch in der Lage sein, in Textdokumenten Konzepte und Zusammenhänge zu erkennen, Fakten zu extrahieren und Zusammenfassungen automatisch zu erstellen. Diese Fähigkeiten sind gegenüber existierenden Erkennungsverfahren ein großer Fortschritt auf dem Weg zu einem effizienteren Informationsmanagement. Auch für die Analyse und Ordnung von multimedialen Inhalten sollen diese Technologien in Zukunft genutzt werden.

Grundlage für dieses ambitionierte Vorhaben ist die Weiterentwicklung von semantischen Technologien, mit denen ein Computer Inhalt und Bedeutungen von Daten automatisch erkennen und in einen Kontext setzen kann. Das ist besonders bei der Ordnung und Auswertung von unstrukturierten Daten wichtig. Ein weiterer Arbeitsschwerpunkt von ORDO ist die Entwicklung effizienter Verfahren zum Aufbau von Ordnungssystemen, mit denen Daten modelliert, verwaltet und organisiert werden können. Durch innovative linguistische und statistische Methoden sollen diese Ontologien noch effektiver eingesetzt und zum Teil automatisiert erstellt werden – ein entscheidender Schritt, um im Internet besser für Ordnung zu sorgen.

Ziele von ORDO

- Informationen semantisch ordnen und priorisieren
- Konzepte und Zusammenhänge in Texten erkennen, Fakten extrahieren, Zusammenfassungen automatisch erstellen
- Zusammenhänge von Informationen visualisieren

6
PROCESSUS – Wissen in Unternehmen besser nutzen

Im globalen Wettbewerb um Innovationen stehen Unternehmen vor der Herausforderung, ihr Wissen optimal zu nutzen. Im Rahmen von PROCESSUS wird eine Lösung entwickelt, damit die richtigen Informationen zur richtigen Zeit beim richtigen Mitarbeiter ankommen. So können Kosten gesenkt und der betriebliche Ablauf kann verbessert werden.

Wissen ist eine elementare Grundlage für die Wettbewerbsfähigkeit der deutschen Industrie. Aber viele entscheidende Informationen verbergen sich in internen Dokumenten – in Texten, Verträgen, Analysen, Plänen, Protokollen und E-Mails. Das wertvolle Wissen, das in diesen Dokumenten steckt, ist den Mitarbeitern daher häufig nur bedingt zugänglich.

Herkömmliche Erfassungssysteme können die Dokumente zwar analysieren, aber nicht in einen übergeordneten Zusammenhang bringen. Die Datenquellen werden bisher nur nach ihrem Inhalt beurteilt und dann in verschiedenen Datenbanken für die jeweiligen Zielgruppen abgespeichert. Die Systeme erfassen noch nicht den übergeordneten Unternehmensprozess, in dem eine Information verwendet werden kann. Hier setzt PROCESSUS an – und versorgt jeden Mitarbeiter gezielt mit dem für ihn gerade relevanten Wissen.

Die Herausforderungen für PROCESSUS sind vielfältig. Die herkömmlichen Erfassungssysteme müssen um die Fähigkeit zur Analyse dynamischer Arbeitsabläufe erweitert werden. Unterschiedliche Datenbanken müssen miteinander vernetzt werden, damit es für den Nutzer keine Rolle mehr spielt, wo die Informationen abgelegt sind. Verschiedene Fachterminologien müssen durch ein Computerprogramm mit Hilfe semantischer Technologien analysiert und harmonisiert werden. Um passgenaue Informationen für den Nutzer zu liefern, spielt die Analyse der Arbeitsprozesse ebenfalls eine wichtige Rolle. Sobald PROCESSUS erkennt, in welchem Kontext eine Information benötigt wird, wird sie an diesen angepasst.

Praxistest in Maschinenbau und Informationstechnologie

Die im Rahmen von PROCESSUS erstellten Konzepte und Lösungen werden im Maschinenbau und in der Informationstechnologie beispielhaft erprobt. Im

Maschinenbau liegt der Fokus auf Anwendungen in der Antriebs- und Automatisierungstechnik. Gerade der Maschinenbau weist hochkomplexe Geschäftsprozesse und sehr unterschiedliche Nutzergruppen auf. Eine typische Prozesskette besteht aus Vorentwicklung, Konstruktion, Produktion, Verkauf und Kundendienst. In diesen Prozessschritten wird dasselbe Produkt in sehr unterschiedlichen Bezügen und Kontexten bearbeitet. So beschäftigen sich Spezialisten in der Konstruktion meist mit technischen Informationen, deren Kollegen im Vertrieb eher mit Markt- und Wettbewerbsanalysen. Dadurch wird es notwendig, vielfältige Datenbestände bereitzustellen – und zwar im Kontext des jeweiligen Prozessschritts.

PROCESSUS wird in solchen Konstellationen eine neue Art des Zugangs zu Informationen ermöglichen: Der Nutzer beschreibt sein Problem und der Computer bezieht bei der Bearbeitung der Anfrage andere, im selben Kontext bereits erfolgreich umgesetzte Lösungen mit ein. So können dem Nutzer auch neue Aspekte und Möglichkeiten aufgezeigt werden, etwa für den Einsatz von bestehenden Produkten in branchenfremden Märkten. Denkbar ist zum Beispiel, dass einem Vertriebsmitarbeiter bei einem Hersteller von Anlagen zur Produktion von Lacken und Farben durch eine auf PROCESSUS gestützte Recherche auffällt, wie leicht sich ein Bauteil, beispielsweise ein Greifarm oder ein Etikettiergerät, auch bei der Herstellung von Getränkeabfüllanlagen einsetzen lässt.

Diese Erkenntnisse lassen sich ebenso auf den Bereich der Informationstechnologie bei der Verknüpfung unterschiedlichster Computeranwendungen, den sogenannten serviceorientierten Architekturen, übertragen. Für die Softwareindustrie wird im Rahmen von PROCESSUS dabei eine Infrastruktur zur Verwaltung von Inhalten entwickelt, um Softwarekomponenten nach ihren Fähigkeiten zur Lösung von Problemen zu klassifizieren und nicht nach ihren technologischen Merkmalen. Aus diesem Werkzeugkasten lassen sich dank der durch PROCESSUS beschriebenen Inhalte die passenden Teile für eine Anwendung einfach kombinieren.

Diese beiden Anwendungsbeispiele zeigen, wie es die im Rahmen von PROCESSUS entwickelten Technologien erlauben, Branchen- und Prozesswissen schneller zu verknüpfen und neue Produkte, Lösungen und Geschäftsmodelle zu entwickeln. Damit verschaffen sie Unternehmen einen Wettbewerbsvorteil: Sie werden schneller, weil für Mitarbeiter Wissen immer dort verfügbar ist, wo es gebraucht wird, und sie sparen Recherche- und Entwicklungskosten. Unternehmen gewinnen darüber hinaus einen Qualitätsvorsprung, weil das bereitgestellte Wissen besonders hochwertig ist. So trägt PROCESSUS dazu bei, dass Unternehmen ihr Wissen optimal nutzen, um im globalen Wettbewerb besser bestehen zu können.

Synergie mit TEXO

Obwohl die Ergebnisse von PROCESSUS auch alleine nutzbar sind, bietet insbesondere die Kombination mit den im Anwendungsszenario TEXO entwickelten Technologien großes Potenzial. Durch ihren modularen Aufbau können die Komponenten von PROCESSUS als eigenständige Dienste auf einer TEXO-Plattform genutzt werden. Damit werden die von PROCESSUS bereitgestellten Informationen über TEXO individuell und kontextbezogen abrufbar und das Wissen in Unternehmen kann noch besser genutzt werden.

Ziele von PROCESSUS

- Herkömmliche Content-Management-Ansätze um Prozessorientierung erweitern
- Zugang zu relevantem Unternehmenswissen vereinfachen
- Kosten reduzieren und Effizienz steigern durch semantische Technologien

7
TEXO – Infrastruktur für internetbasierte Dienste

Während der Handel mit Inhalten wie Musik und Videos im Internet bereits alltäglich ist, werden Dienstleistungen im Netz bisher noch nicht in diesem Umfang genutzt. Im TEXO-Anwendungsszenario wird deshalb eine Infrastruktur entwickelt, mit der sich Dienstleistungen im Internet einfacher kombinieren und in Anspruch nehmen lassen. Ein wichtiger Schritt auf dem Weg zum Internet der Dienste.

Viele Verbraucher und Unternehmen sehen sich einem Problem gegenüber: Die Suche nach passenden Dienstleistungen im Internet ist langwierig und aufwändig. Erschwerend kommt hinzu, dass bei komplexen Vorgängen oft unterschiedliche Ansprechpartner für verschiedene Teilleistungen zuständig sind und sich diese nicht sinnvoll kombinieren lassen. Wer beispielsweise aus dem Ausland ein Geschäft in Deutschland eröffnen möchte, merkt schnell, wie aufwändig die Unternehmensgründung ist: Standortwahl, Handelsregistereintrag, Marktanalyse und die Einstellung geeigneter Mitarbeiter verschlingen Zeit und Ressourcen. Gerade für viele kleinere und mittlere Unternehmen wäre es eine große Erleichterung, wenn sich alle diese Schritte über einen zentralen Ansprechpartner abwickeln ließen. Doch welche Anbieter haben alle notwendigen Dienstleistungen im Programm? Und wie findet man sie?

Grundlage für eine innovative Dienstleistungswirtschaft

TEXO entwickelt eine integrierte Online-Plattform, die diese Probleme lösen kann. Damit wird eine wichtige Grundlage für eine neue Dienstleistungswirtschaft im Internet gelegt, in der Serviceleistungen wie Güter handelbar sind und zu neuen Angeboten kombiniert werden können. Die Plattform ermöglicht den Handel von technischen Services und von konventionellen Angeboten wie Beratungsdienstleistungen und trägt dabei sowohl den Interessen der Anbieter als auch den Bedürfnissen der Kunden Rechnung.

Vielfältiger Nutzen für Anbieter

Vor allem für kleine und mittlere Unternehmen eröffnen sich durch die Möglichkeit, Dienstleistungen über das Internet anzubieten, große Chancen. Sie können ihre Dienste weltweit präsentieren und neue Kunden gewinnen. Auch können Unternehmen das eigene Angebot durch Kombination mit Diensten anderer Anbieter erweitern und so mit innovativen

Servicepaketen neue Märkte erschließen. TEXO schafft die technologischen Voraussetzungen, damit neue Dienste schneller entwickelt und Angebote von Partnern einfacher integriert werden können. Dadurch können sich vor allem kleine und mittlere Unternehmen auf ihre Kompetenzen konzentrieren sowie Zeit und Geld sparen – denn Aufgaben, die nicht zum Kerngeschäft gehören, können einfach an externe Dienstleister ausgelagert werden.

Neue Möglichkeiten für Kunden

Sowohl Verbraucher als auch Unternehmen können als Kunden von der neuen Dienstleistungsplattform profitieren. Die drei folgenden Beispiele verdeutlichen den Nutzen von TEXO.

Beispiel 1: Der „Öko-Kalkulator“ für Unternehmen

Ein Unternehmen benötigt für den Export seiner Produkte ein Zertifikat, das die Recyclingquote dieser Produkte dokumentiert. Nur ausgewählte Anbieter dürfen solche Zertifikate erstellen. Auf der TEXO-Plattform wird das Unternehmen schnell fündig: Dort werden verschiedene „Öko-Kalkulatoren“ als Dienste für die Zertifizierung angezeigt. Der ausgewählte „Öko-Kalkulator“ wird automatisch in die Arbeitsumgebung des Unternehmens integriert und kann dort für die Berechnung der Recyclingquote von Produkten als Voraussetzung für die Zertifizierung genutzt werden. So können Unternehmen dank TEXO die für sie passende Dienstleistung ohne aufwändige Recherche finden und über eine einfach bedienbare Plattform in Anspruch nehmen.

Beispiel 2: Zentrale Abwicklung von Behördengängen

Wenn Kinder zur Welt kommen, sehen sich die Eltern vielen Aufgaben gegenüber: Behördengängen, Pflichtuntersuchungen, E-Mails an Freunde und vielem mehr. Mit der TEXO-Plattform lassen sich die dafür notwendigen Dienste von Unternehmen und Behörden bündeln und auf einer einzigen Website anbieten. Damit wird es für die Eltern möglich, alle anstehenden Aufgaben zentral über eine intuitive Benutzeroberfläche zu erledigen. So müssen die Eltern zum Beispiel nur einmal ihren Namen und ihre Adresse angeben und erhalten automatisch sowohl Informationen zu Kinderärzten in ihrer Umgebung wie auch alle wichtigen Formulare der zuständigen Ämter.

Beispiel 3: Vereinfachte Unternehmensgründung

Die Dienstleistungsrichtlinie der Europäischen Union soll EU-Bürgern die Unternehmensgründung erleichtern, indem künftig alle im Zusammenhang mit dem Gründungsprozess notwendigen Behördengänge über einen Ansprechpartner abgewickelt werden können. Die TEXO-Plattform ermöglicht eine flexible Auswahl und Kombination der Dienste verschiedener Behörden, um für die jeweiligen Gründer passgenaue Pakete zu schnüren und den aktuellen Bearbeitungsstand zu verfolgen. Im Rahmen von TEXO wird dabei eine solche individuelle und dynamische Zusammenstellung von Diensten für den speziellen Anwendungsfall des Jungunternehmers entwickelt, der ein Geschäft eröffnen möchte.

SOA als Grundlage

Die technologische Grundlage von TEXO ist die sogenannte Serviceorientierte Architektur (SOA). SOA ist eine Technik in der Informationstechnologie, mit der verschiedene Computeranwendungen zu einer einzigen Anwendung zusammengefasst werden können. Die Forscher von TEXO bauen auf SOA auf und entwickeln eine Infrastruktur, die mit semantischen Technologien die Bedeutung von Informationen und Services so beschreibt, dass ein Computer diese versteht. Damit wird die Grundlage für eine neue Dienstleistungswirtschaft im Internet geschaffen.

Ziele von TEXO

- Dienstleistungsplattform im Internet schaffen
- Serviceleistungen sollen wie Güter handelbar sein und zu neuen Angeboten kombiniert werden
- Die Grundlage für eine offene und sichere Umgebung für den Handel mit Diensten im Internet bereitstellen
- Neue Geschäftsmodelle, insbesondere für KMU

8
THESEUS-Basistechnologien

Das Internet enthält mehr Wissen als alle Archive und Datenbanken in der Geschichte der Menschheit. Um dieses Wissen in Zukunft besser und einfacher zugänglich zu machen, werden im Rahmen von THESEUS die notwendigen Basistechnologien und Standards entwickelt.

Im Fokus von THESEUS steht die Entwicklung semantischer Technologien, mit denen die inhaltliche Bedeutung von Informationen erfasst und ausgewertet werden kann. Dadurch können Computerprogramme Informationen wie Texte, Bilder oder Ton- und Videoaufzeichnungen automatisch inhaltlich analysieren, verknüpfen und daraus logische Schlüsse ziehen – eine wichtige Weiterentwicklung auf dem Weg zum Internet der Dienste.

Um diese Evolution des Internets zu ermöglichen, müssen innovative Basistechnologien und technische Standards entwickelt werden. Die im Rahmen von THESEUS erforschten Basistechnologien werden von den Industriepartnern im THESEUS-Konsortium in sechs Anwendungsszenarien entwickelt und erprobt. Dabei soll auch geprüft werden, wie diese Technologien möglichst schnell für neue Werkzeuge, Dienste und Geschäftsmodelle im Internet genutzt werden können.

Die Basistechnologien im Überblick:

1. Automatische Erzeugung von Metadaten

Die THESEUS-Forschungspartner entwickeln neue Methoden, um Medieninhalte zu erfassen. Diese Methoden basieren auf Metadaten. Das sind Informationen zu einem Objekt, die in einer Datenbank registriert werden und bei einer Suche den Weg zu diesem Objekt weisen – zum Beispiel der Name eines Buchautors oder die Epoche, in der ein Film spielt. Mit den neuen Verfahren können Metadaten für verschiedene Medieninhalte erstellt werden, also unter anderem für Texte, Fotos, Audio- und Videodateien. Ein Entwicklungsschwerpunkt liegt dabei auf Ansätzen zur automatischen Generierung von Metadaten. Einen weiteren Schwerpunkt bildet der Einsatz von semantischen Technologien, um den Kontext eines Medieninhalts zu erfassen. Ziel

ist es, thematisch ähnliche Inhalte aus unterschiedlichen Quellen zu einer Gruppe zusammenzufassen, die dem Nutzer möglichst viele Informationen bietet.

2. Schnelle Verarbeitung multimedialer Dokumente

Damit Nutzer bei einer Recherche in komplexen multimedialen Datenbanken nicht zu lange auf ein Ergebnis warten müssen, werden im Rahmen von THESEUS hocheffiziente Algorithmen für Metadaten entwickelt. Sie erlauben es, auch in Datenbanken mit mehreren hunderttausend Objekten die Inhalte von Bildern und Videos schnell zu durchsuchen. Ein weiterer Schwerpunkt ist die Komprimierung von Daten, insbesondere von Bilddateien. Da die Qualität und die Auflösung von digitalem Bildmaterial weiter ansteigen werden, werden neue Kodierverfahren benötigt, um die Informationen platzsparend verpacken und speichern zu können. Um die Bildsuche zu verbessern, entwickeln die Forscher zudem Bilderkennungssysteme. Sie werden den Computer in die Lage versetzen, den Gegenstand selbstständig zu erfassen, der auf einem Foto oder in einem Video dargestellt wird.

3. Innovatives Ontologien-Management

Um Informationen effizient analysieren und verknüpfen zu können, werden im THESEUS-Forschungsprogramm neue Methoden und Werkzeuge für das Management von Ontologien entwickelt. Ontologien sind Ordnungssysteme, die den Inhalt von Daten näher beschreiben und Aufschluss über ihren Zusammenhang geben. Ziel der Entwickler ist es, möglichst aussagekräftige Ontologien zu schaffen – bis hin zu automatisierten logischen Schlussfolgerungen (Reasoning), die ein Programm selber treffen kann. Ähnlich wie mit Metadaten können so verschiedene Wissensbestände verknüpft werden, um schnellere und zielführendere Recherchen zu ermöglichen.

4. Maschinelles Lernen

Ein weiteres Projekt von THESEUS ist die Entwicklung skalierbarer Technologien, mit denen Metadaten strukturiert und Ontologien modelliert werden können. Diese Technologien erlauben es Anwendungen, effizient mit den Ressourcen eines Computers umzugehen, etwa um so viele Rechenschritte in möglichst kurzer Zeit zu erledigen. Hierzu werden innovative Verfahren zur Ermittlung von Beziehungen in Texten, Bildern sowie Audio- und Videodateien verwendet, die auf den Methoden des maschinellen Lernens basieren. Beim maschinellen Lernen verarbeitet ein Computer nicht nur Informationen, sondern erkennt auch Zusammenhänge und Gesetzmäßigkeiten zwischen ihnen. Dieses erworbene Wissen kann er dann selbstständig auf andere Daten übertragen.

5. Situationsbewusste Dialogverarbeitung

Damit der Computer im Auftrag des Nutzers handeln kann, muss er dessen Anliegen verstehen. Für diesen Dialog zwischen Mensch und Maschine werden im Rahmen von THESEUS neue Funktionen entwickelt, die sich flexibel in verschiedenen Anwendungen einsetzen lassen. Je nachdem, wie und wofür der Nutzer mit dem Computer kommunizieren will, kann eine eigene Kommunikationsform gewählt werden. So ermöglichen innovative Algorithmen multimodale Benutzeroberflächen, die durch Sprache und Gestik bedient werden können. Mit solchen Schnittstellen können Nutzer ihre Anfragen intuitiv formulieren und sie im gesprochenen Dialog mit dem System verfeinern. Eine spezielle Komponente, der semantische Mediator, fungiert dabei innerhalb des Computers als Schnittstelle zwischen der multimodalen Benutzeroberfläche und den verschiedenen Metadatenquellen. Er wandelt eine gesprochene Anfrage in einen semantisch korrekten Datensatz um, der für eine Suche im System nötig ist.

6. Zukunftweisende Benutzeroberflächen

Um Zusammenhänge zwischen verschiedenen Daten, Metadaten und Dokumenten einfacher erkennbar zu machen, werden zudem im THESEUS-Programm neue grafische Benutzeroberflächen entwickelt. So können zum Beispiel die Ergebnisse von Suchanfragen als sogenannte Wissensnetze dargestellt werden, die zeigen, wie die Ergebnisse sowohl mit dem Suchbegriff als auch untereinander zusammenhängen. Diese Darstellungsform

ermöglicht es, einen einfachen Überblick über ein Themengebiet zu erhalten und die benötigten Informationen schneller zu finden. Davon profitieren nicht nur private Nutzer, sondern auch Unternehmen, die durch definierte Standards solche Benutzeroberflächen einfach integrieren und an ihre spezifischen Anforderungen anpassen können – und so Entwicklungskosten sparen.

7. Neue Wege zur Identifizierung von Inhalten

Die THESEUS-Forschungspartner erarbeiten zudem neue Ansätze, mit denen Inhalte eindeutig identifiziert werden können, selbst wenn diese leichten Veränderungen unterzogen wurden. So lassen sich zum Beispiel Versionen eines Films im Produktionsprozess verfolgen. Aber auch Urheber- und Vermarktungsrechte an geistigem Eigentum lassen sich schützen. So kann erreicht werden, dass trotz der durch das Internet stark erweiterten Verfügbarkeit von multimedialen Inhalten die Rechte der Urheber nicht verletzt werden.

8. Evaluation der Basistechnologien

Die im Rahmen von THESEUS entwickelten Basistechnologien werden von Experten hinsichtlich ihrer Qualität beurteilt. So werden beispielsweise neue Technologien zur Sprach- und Bilderkennung oder zur automatischen Zuordnung von Metadaten auf ihre Zuverlässigkeit, Funktionalität und Tauglichkeit geprüft. Die Ergebnisse dieser Evaluation fließen dann wieder in den Forschungs- und Entwicklungsprozess ein und tragen dazu bei, die Ergebnisse zu optimieren. Die Bewertungen und Expertengutachten dienen darüber hinaus der Entwicklung von technischen Standards.

9
THESEUS-Begleitforschung

Rund um den Globus gibt es viele Forschungsanstrengungen, das Internet weiterzuentwickeln. Eine partnerschaftliche Brücke zu diesen Projekten schlägt die Begleitforschung, ein THESEUS-Modul mit internationaler Ausrichtung. Es sucht interessante Projekte, knüpft Kontakte zu den Forschern und baut so ein weltweites Netzwerk auf, von dem alle Wissenschaftler profitieren.

Die THESEUS-Begleitforschung erfolgt im Auftrag des BMWi und unterstützt die THESEUS-Forschungsvorhaben bei deren internationaler Vernetzung. Aufbau und Betrieb der Begleitforschung liegen in den Händen des Fraunhofer-Instituts für Software- und Systemtechnik ISST und der Beratungsagentur ESG Consulting.

Aufgabe der Begleitforschung ist es, die in THESEUS entwickelten Technologien und Anwendungen bzw. Geschäftsmodelle vergleichbaren Initiativen im europäischen und internationalen Umfeld gegenüberzustellen. Sie führt vergleichende Untersuchungen durch, um Gemeinsamkeiten und Unterschiede zwischen den THESEUS-Vorhaben und anderen Vorhaben zu den im THESEUS-Programm bearbeiteten Forschungsthemen aufzuzeigen. Die Ergebnisse dienen dem BMWi und dem THESEUS-Konsortium dazu, strategische Rückschlüsse für eine Positionierung und Weiterentwicklung des THESEUS-Programms ableiten zu können. Dazu wird die THESEUS-Begleitforschung die weltweiten Überlegungen und Vorhaben zur Entwicklung des Internets der Zukunft besonders berücksichtigen. Die in diesen Arbeiten verfolgten Ansätze zur Entwicklung neuer Angebote für Informationen und Dienste zur Weiterentwicklung des Internets zu einer internetbasierten Wissensinfrastruktur werden mit besonderer Aufmerksamkeit verfolgt.

Die Metapher]WiSSEN.NUtZEN[, die die THESEUS-Begleitforschung ihren Arbeiten voranstellt, charakterisiert die mit internetbasierten Wissensinfrastrukturen verbundenen Erwartungen und die dazu benötigten technischen Lösungen.

Die Begleitforschung erfasst Nutzerdienste und Funktionalitäten, die im Internet der Dienste bereitgestellt werden sollen. Dazu wird auch die Bedeutung semantischer Technologien für die erfolgreiche Einführung und Nutzung des Internets der Dienste abgeschätzt. In der Entwicklung des Internets der Zukunft entwickelte und genutzte Engineering-Konzepte und -Verfahren werden von der Begleitforschung, soweit zugänglich, vorgestellt und zur Diskussion gestellt.

Gleichzeitig soll durch die Vernetzung mit wirtschaftlichen und gesellschaftlichen Akteuren in Deutschland der Wissenstransfer in verschiedene Anwendungsbereiche aus THESEUS heraus angestoßen werden. Dazu werden öffentlichkeitswirksame Veranstaltungen angeboten, Publikationen veröffentlicht und die Ergebnisse der Begleitforschung über das Internet zugänglich gemacht.

10
THESEUS Talente 2008 – Ideenwettbewerb zur Nachwuchsförderung

In einem Ideenwettbewerb suchte das THESEUS-Konsortium nach klugen Köpfen. Nachwuchstalente erhielten im Rahmen von „THESEUS Talente 2008“ Gelegenheit, die Forschungsarbeit im THESEUS-Programm mit eigenen Ideen aktiv mitzugestalten.

Nachwuchswissenschaftler, Studierende, Schüler und junge Entwickler – sie alle waren erstmals aufgerufen, Ideen für die Nutzung von semantischen Technologien im Rahmen von THESEUS einzureichen und so an der Entwicklung des Internets der Dienste mitzuarbeiten. Zur Bearbeitung standen dabei fast 30 Themenbereiche rund um die Basistechnologien und Anwendungsszenarien von THESEUS zur Auswahl.

Ob Verfahren zur Bildklassifizierung, die Bilder auf Grundlage von definierten Merkmalen automatisch kategorisieren, oder Vorschläge für neue Dienstleistungen im Internet – die möglichen Denkansätze waren so vielfältig wie das THESEUS-Forschungsprogramm selbst. Die Juroren konnten aus mehr als 150 Einreichungen auswählen und bewerteten diese nach den Kriterien Qualität, Originalität, Innovationsgrad sowie Umsetzbarkeit.

Den ersten Platz belegte die angehende Sprachwissenschaftlerin Sonja Kraus aus Grevenbroich, die die Jury mit ihrem Beitrag „Semantstrategien“ überzeugte. Die Studentin beschäftigt sich mit der Anreicherung von Inhalten durch Metadaten. Dieses sogenannte Tagging verbessert die Suche und Vernetzung von Inhalten. Sonja Kraus hat ein leicht zu bedienendes Werkzeug entwickelt, das Nutzern automatisch generierte Vorschläge unterbreitet, mit denen sie Inhalten im Internet Schlagworte zuweisen können. Damit wird die Zuweisung von Metadaten einfacher und schneller – ein wichtiger Schritt auf dem Weg zum Internet der Zukunft.

Geldpreise und Patenschaften

Insgesamt wurden Prämien in Höhe von 50.000 Euro ausgezahlt. Neben den Geldpreisen bot man den 14 Preisträgern die Möglichkeit, ihre Ideen im Rahmen von Praktika bei Unternehmen und Forschungseinrichtungen des THESEUS-Konsortiums umzusetzen. So können die Nachwuchstalente die Zusammenarbeit in einem Netzwerk namhafter Partner kennen lernen und werden dabei von erfahrenen Experten betreut.

Der Wettbewerb THESEUS Talente, der vom Bundesministerium für Wirtschaft und Technologie in Zusammenarbeit mit dem THESEUS-Konsortium durchgeführt wurde, ermöglichte so, den IT-Nachwuchs gezielt zu fördern und das Forschungsprogramm für neue Ideen zu öffnen.

THESEUS Talente – die Fakten

- Ideenwettbewerb für den wissenschaftlichen Nachwuchs in Deutschland
- Aufgaben anhand konkreter Fragestellungen aus über 30 Bereichen zu den Basistechnologien und aus den Anwendungsbereichen
- Insgesamt 50.000 Euro Preisgeld
- Zusätzliche Patenschaften durch die Unternehmen des THESEUS-Konsortiums für die Preisträger

11
THESEUS Mittelstand 2009 – Technologiewettbewerb für den Mittelstand

In vielen Technologiebereichen nehmen mittelständische Unternehmen eine Vorreiterrolle ein. Denn sie sind bei der Erschließung neuer Märkte besonders schnell und flexibel. Dabei haben aber gerade kleine und mittlere Unternehmen häufig nur wenig Zeit und knappe Ressourcen für Forschung und Entwicklung. Umso mehr profitieren sie von den Kooperationsmöglichkeiten des Förderwettbewerbs „THESEUS Mittelstand 2009“.

Mit dem Technologiewettbewerb „THESEUS Mittelstand 2009“ erhalten kleine und mittlere Unternehmen aus Deutschland die Möglichkeit, die im Rahmen von THESEUS entwickelten Technologien zu erproben und in neue Produkte und Verfahren umzusetzen. Der Schwerpunkt liegt dabei auf einer hohen Anwendungsorientierung. Der Wettbewerb wird vom Bundesministerium für Wirtschaft und Technologie in enger Zusammenarbeit mit dem THESEUS-Konsortium durchgeführt. Für den Wettbewerb stellt das BMWi Fördermittel in Höhe von rund 10 Mio. Euro für eine Laufzeit von bis zu drei Jahren zur Verfügung.

Innovative Anwendungen und Verfahren gesucht

Die Ergebnisse aus der ersten Programmphase, die auf der Entwicklung von THESEUS-Basistechnologien und deren Erprobung in den Anwendungsszenarien beruhen, bilden die Basis für die Einbindung der kleinen und mittleren Unternehmen. Auf Grundlage der vorliegenden Forschungsergebnisse haben die Unternehmen Gelegenheit, neue Technologien, Produkte und Verfahren zu entwickeln und zu erproben. Gefördert werden anwendungsorientierte Projekte wie die Entwicklung neuer Dienste und Geschäftsmodelle. Bewerben können sich mittelständische Unternehmen mit bis zu 1.000 Beschäftigten. Diese können ihre Projekte einzeln oder im Verbund mit anderen Partnern einreichen. Um passende Partner zu finden, steht interessierten Unternehmen im Internet ein Forum zur Verfügung.

Der Wettbewerb sieht ein zweistufiges Verfahren vor: In der Bewerbungsphase reichen interessierte Unternehmen bis zum 1. Dezember 2008 ihre Projektskizzen ein. Diese werden von einer unabhängigen Jury im Hinblick auf Originalität, Umsetzbarkeit sowie Verwertungschancen und Nachhaltigkeit bewertet. In dieser Bewertung werden auch die Kompetenz der Bewerber und die Relevanz des Beitrags für die Programmziele von THESEUS mit einbezogen. Den von der Jury ausgewählten Bewerbern wird dann in einer anschließenden Konsultationsphase Gelegenheit gegeben, im Rahmen von Workshops mit dem THESEUS-Konsortium die eingereichten Projektskizzen weiter auszuarbeiten und auf dieser Basis die Förderanträge zu stellen.

Mehr Wachstum und Beschäftigung.

„THESEUS Mittelstand 2009“ ist ein wichtiger Beitrag zur Förderung von Spitzenforschung im Mittelstand. Der frühzeitige Transfer der wissenschaftlichen Ergebnisse aus THESEUS in die mittelständische Wirtschaft verbessert die Wettbewerbsfähigkeit der Unternehmen und stärkt den Multiplikatoreffekt – die Ergebnisse können auch von weiteren Unternehmen schneller genutzt werden. So werden wichtige Impulse für mehr Innovation, Wachstum und Beschäftigung am Standort Deutschland gesetzt.

Mehr Informationen zum Wettbewerb unter:
http://theseus-programm.de/mittelstand

Die Ziele von „THESEUS Mittelstand“

- Öffnung des Programms für eine stärkere Beteiligung des Mittelstands
- Frühzeitiger Transfer von wissenschaftlichen Erkenntnissen in die mittelständische Wirtschaft
- Stärkere Partizipation von kleinen und mittelständischen Unternehmen an den Ergebnissen von THESEUS
- Erweiterung des THESEUS-Programms durch die Entwicklung und Erprobung innovativer Technologien für neue Produkte und Verfahren
- Stärkung von Wirkung und Nachhaltigkeit des THESEUS-Programms

12
Die Partner des THESEUS-Programms

- Deutsche Nationalbibliothek, Frankfurt am Main
- Deutsche Thomson oHG, Hannover
- Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI) GmbH, Kaiserslautern
- empolis GmbH, Gütersloh
- Festo AG & Co. KG, Esslingen
- Forschungszentrum Informatik (FZI), Karlsruhe
- Fraunhofer-Gesellschaft (9 Institute)
- Institut für Rundfunktechnik GmbH (IRT), München
- intelligent views GmbH, Darmstadt
- Ludwig-Maximilians-Universität (LMU), München
- Lycos Europe GmbH, Gütersloh
- mufin GmbH, Berlin
- moresophy GmbH, München
- ontoprise GmbH, Karlsruhe
- SAP AG, Walldorf
- Siemens AG, München
- Technische Universität München, Garching
- Technische Universität Darmstadt, Darmstadt
- Technische Universität Dresden, Dresden
- Universitätsklinikum Erlangen – Radiologisches Institut, Erlangen
- Universität Karlsruhe, Karlsruhe
- VDMA Gesellschaft für Forschung und Innovation (VFI), Frankfurt am Main

Diese Druckschrift wird im Rahmen der Öffentlichkeitsarbeit des Bundesministeriums für Wirtschaft und Technologie herausgegeben. Sie wird kostenlos abgegeben und ist nicht zum Verkauf bestimmt. Sie darf weder von Parteien noch von Wahlwerberinnen/Wahlwerbern oder Wahlhelferinnen/Wahlhelfern während eines Wahlkampfes zum Zweck der Wahlwerbung verwendet werden. Missbräuchlich ist insbesondere die Verteilung auf Wahlveranstaltungen und an Informationsständen der Parteien sowie das Einlegen, Aufdrucken oder Aufkleben parteipolitischer Informationen oder Werbemittel. Untersagt ist gleichfalls die Weitergabe an Dritte zum Zwecke der Wahlwerbung. Unabhängig davon, wann, auf welchem Weg und in welcher Anzahl diese Schrift der Empfängerin/dem Empfänger zugegangen ist, darf sie auch ohne zeitlichen Bezug zu einer bevorstehenden Wahl nicht in einer Weise verwendet werden, die als Parteinahme der Bundesregierung zugunsten einzelner politischer Gruppen verstanden werden könnte.