Kaum eine andere Technologie hat das Alltagsleben und den Umgang mit Informationen so schnell und gravierend verändert wie das World Wide Web (WWW). Durch die rasante Entwicklung des Web wandelte sich die Industriegesellschaft allmählich zu einer Informationsgesellschaft12. Das Web veränderte auch die Art und Weise, wie die Menschen miteinander kommunizieren, wie Informationen verbreitet und wiedergefunden werden und es beeinflusste ebenfalls die Ausführung von Geschäften bzw. Geschäftsbeziehungen in Unternehmen13. Dank einer Standardisierung des Web14 und der darauf aufbauenden Infrastruktur zeichnet sich das Web durch eine Reihe von erheblichen Vorteile aus. Folgende Merkmale sind für das Web prägend: „Aktualität und Verfügbarkeit“15. Unter Aktualität versteht man, dass Informationen zu jeder Zeit auf den neuesten Stand gebracht werden und durch die Verfügbarkeit ist u. a. die Lokalität von Informationen bzw. Informationsquellen kaum von Belang, da man auf eine beliebige bereitgestellte Information irgendwo im Web von jedem Ort aus zugreifen kann. Diese Tatsache wird auch als universelle Verfügbarkeit bezeichnet16. Dadurch stellt sich der Weg zur Informationsbeschaffung weniger aufwändig und kostengünstiger dar und darüber hinaus sind die Informationen für eine breitere Bevölkerungsschicht öffentlich zugänglich. Die zunehmende Bedeutung des Web sowohl in kommerzieller als auch in gesellschaftlicher Hinsicht führte zu einer Liberalisierung der Bereitstellung von Informationen. Vor dem Zeitalter des WWW gestaltete sich die Verbreitung und Veröffentlichung von Informationen für das breite Publikum als ausgesprochen schwierig, da Informationen nur einigen bestimmten Gruppen, genauer gesagt den sogenannten „Informationsoligopolisten“ vorbehalten waren. Heutzutage werden solche Schranken durch das Web umgangen es bietet einzelnen Personen oder auch kleineren Interessengruppen die Möglichkeit, ihre Informationen dem breiten Publikum zur Verfügung zu stellen, ohne irgendwelche Kosten dafür aufbringen zu müssen. All Diese Vorzüge und Vorteile des Web spiegeln die Ideen des WWW Erfinders Tim Berners-Lee wider. In seinem „WorldWideWeb: Summary“18 beschreibt Tim Berners-Lee das Web folgendermaßen: „The WWW project merges the techniques of information retrieval and hypertext to make an easy but powerful global information system“19. Dies bedeutet: Das WWW Projekt wird mit den Techniken des Information Retrieval und den Ideen des Hypertext zu einem globalen Informationssystem vereinigt bzw. vermischt. Die Betonung liegt hier auf dem Begriff „Global Information System“, also ein System, in dem Informationen gespeichert, wiedergefunden und einer breiten Masse zur Verfügung gestellt werden. Ein anderer Pionier des WWW, Kevin Hughes, beschreibt das Web in seinem Paper20 noch vortrefflicher: „The WorldWideWeb (W3) is a wide-area hypermedia information retrieval initiative aiming to give universal access to a large universe of documents“21. Kevin Hughes interpretiert das WWW als eine Art universelles Dokumentenarchiv, auf das immer und zu jeder Zeit zugegriffen werden kann. D. h. jeder Benutzer eines Computernetzwerks ist in der Lage, ständig auf eine Reihe von Medien zurückgreifen zu können. Heutzutage ist die Benutzung des WWW kaum aus dem Alltagsleben wegzudenken. Aktuell existieren diverse kommerzielle Nutzungsmöglichkeiten des Web wie z. B. das Einkaufsportal www.amazon.de oder auch unabhängige Nachrichtenportale. In letzter Zeit gewann das sogenannte Social Networking durch Portale wie Facebook oder Twitter an Bedeutung. Das Marktstudienportal markt-studie.de ist ein typisches Beispiel für ein Einkaufsportal, also eine kommerzielle Nutzung des Web. Auch das Portal nutzt die Annehmlichkeiten und Vorteile des Web. Die Veröffentlichung der Studien im Studienportal ist mit keinen Kosten verbunden und die Daten werden nach einem bestimmten Monatsrhythmus ständig aktualisiert. Mit Hilfe der Social Networking Portale können die Links zu den Studien verbreitet werden, was dazu führt, dass diese Daten eine stärkere Präsenz im Web erhalten. Dies spiegelt die beiden o. g. prägnanten Merkmale des Web wider: „Aktualität“ und „Verfügbarkeit“.

Trotz aller Vorzüge und Vorteile des Web tauchen doch auch etliche Probleme auf. Man darf die Tatsache nicht ignorieren, dass die im Web vorhandenen Informationen nur von Menschen konsumiert werden und die Lesbarkeit auch nur auf den Menschen beschränkt ist22. Ein gewöhnlicher Web-Content beispielsweise, der aus einer Datenbank erzeugt wird, stellt den Inhalt ohne strukturelle Informationen dar. Maschinen können mit den unstrukturierten Ansammlungen von Daten im Web nichts anfangen, sie können sie nicht weiterverarbeiten. Ein Mensch als Endkonsument der Informationen besitzt die Fähigkeit, auf einer Webseite gefundene Informationen zu interpretieren und sie mit anderen Informationen in Beziehung zu setzen, während auf der anderen Seite die Maschine ohne strukturelle Grundlage der Informationen nicht in der Lage ist, diese Aufgaben zu erfüllen23. So kann ein Mensch im Gegensatz zu der Maschine auf Basis einer Information Schlussfolgerungen ziehen und diese zu einem Gesamtergebnis zusammenfassen. Das Prinzip der Verfügbarkeit, im letzten Kapitel explizit erläutert, bedarf ebenfalls einer näheren Betrachtung. Die Informationen im Web sind zwar prinzipiell überall verfügbar und dem Zugriff darauf sind keine Grenzen gesetzt, dennoch ist es problematisch und schwer, die explizit gesuchten Informationen in der Menge der Informationen, die das Web hergibt, zu finden. Dieses Problem bezieht sich auf Suchmaschinen wie Google oder Yahoo. Das Suchen und Finden von Informationen zu Recherchezwecken ist heutzutage eine typische Webnutzung. Hinter der Suchmaschine verbirgt sich ein ausgefeilter Algorithmus zum Auffinden von Information, letztendlich aber basiert die Suchmethode auf dem Lokalisieren von Zeichenketten im Text24, es handelt sich also bei den gängigen Suchmaschinen um eine stichwortbasierte Suche. Mit der stichwortbasierten Suche erzielt man zwar eine hohe Trefferquote, jedoch sind die Ergebnisse wenig präzise. Aus Erfahrung ist diese Feststellung äußerst zutreffend, da i. d. R. nach einer Suchanfrage bei Google zunächst mehr als Tausende Ergebnisse aufgelistet werden – die meisten dieser Informationen sind nicht relevant und somit kaum nutzbar25. Manchmal kommt es sogar vor, dass überhaupt keine Ergebnisse geliefert werden, so z. B wenn das gesuchte Schlüsselwort in den durchsuchten Web-Dokumenten nicht auftaucht. Diese Feststellung zeigt wie stark vokabularabhängig26 die Suchergebnisse in der stichwortbasierten Suche sind. Eine Alternative bietet die inhaltsbezogene, also semantische Suche. Beispielswei se sollten die Suchergebnisse für den Suchbegriff „Wein“ nicht nur alle Informationen zu Wein liefern, sondern es sollten auch andere alkoholische Getränke oder Weinarten in den Ergebnissen angezeigt werden. Oftmals kommt es aber vor, dass eine exakte Information, nach der ein Nutzer sucht, nicht explizit auf einer einzelne Webseite oder im Web zu finden ist, sondern die kompletten Informationen aus unterschiedlichen, über das Web verteilten Fakten bestehen. Der Nutzer müsste dann die einzelnen Seiten durchsuchen, um die passenden Fakten zusammenzutragen, aus denen er letztlich Schlüsse ziehen kann. Diese Problematik bezeichnet man als Problem des impliziten Wissens. Implizites Wissen wird auch als stilles Wissen bezeichnet und leitet sich aus dem englischen „tacit knowledge“ ab. Nach [RAN01] definiert sich implizites Wissen folgendermaßen: „Es sind Kenntnisse oder Fähigkeiten, die nicht explizit formuliert sind“27. Das Wissen lässt sich nur anzeigen und ist auch nicht erklärbar oder es ist kaum möglich es in Worten auszudrücken28. Beispielsweise findet ein Nutzer erst am Ende seiner Faktensammlung unerwartet eine zusätzliche Information, die sich jedoch als eine gute Ergänzung zur eigentlichen Zielinformation herausstellt. Nicht ausser acht zu lassen ist die Problematik der verschiedenen Kodierungstechniken, Dateiformate, natürlichen Sprachen und der unterschiedliche Aufbau privater Homepages – der unzähligen Webseiten im Web. Die Gründe dafür sind die dezentrale Struktur und Organisation des Web, dies hat zur Folge, dass die vorhandenen Informationen im Web heterogen sind. Dies macht es schwierig, über das Web verteilte Informationen zu sammeln, sie zusammenzufassen und ggf. weiterzuverarbeiten. Als Anwendungsbeispiel könnte man hier den Versuch, die verschiedenen politischen Programme der deutschen Parteien auf Basis ihrer Internetpräsenz miteinander zu vergleichen, nennen. Dieses Vorhaben wird schwer zu bewerkstelligen sein, da eventuell die Webseiten der verschiedenen Parteien unterschiedlich aufgebaut sind. Verfolgt man dieses Beispiel weiter und möchte die Parteiprofile der deutschen und der isländischen konservativen Partei miteinander vergleichen, scheitert dies schon an den unterschiedlichen natürlichen Sprachen, da Island ein anderes Alphabet benutzt. Dies bezeichnet man als das Problem der „Informationsintegration“29. Um diesen verschiedenen Problematiken entgegenzuwirken, entstand beim W3C das Semantic Web Projekt, das auf dem Vorschlag des WWW Erfinders Tim Berners-Lee beruht. Das Konzept und der Vorschlag zum Semantic Web beruhen auf einem Artikel von Tim Berners-Lee, den er gemeinsam mit James Hendler und Ora Lassila verfasst und im amerikanischen Wissenschaftsmagazin „Scientific American“30 veröffentlicht hat. In dem Artikel spricht Tim Berners-Lee die verschiedenen Problematiken des herkömmlichen Web an. Zum einen die Feststellung, dass Informationen im Web einschließlich und nur für den Menschen bestimmt sind und zum anderen, dass Maschinen nicht in der Lage sind, diese Informationen weiterzuverarbeiteten. Die Hauptidee des Semantic Web besteht darin, das homogen verteilte Web aus der Ebene der Darstellung in die Ebene der Daten zu heben. Das Semantic Web soll jedoch das herkömmliche Web nicht ersetzen oder separat zum bestehenden Web existieren, es stellt lediglich es eine Erweiterung des herkömmlichen Web dar, in der Informationen Bedeutungen zugewiesen werden, so dass Maschinen die Informationen „verstehen“ und weiterverarbeitet können. In seinem Artikel erläutert Tim Berners-Lee das Semantic Web am Beispiel von zwei Geschwistern, die einen passenden Arzt für ihre Mutter finden wollen. Wenn geeignete Ärzte gefunden werden, dürfen die Arzttermine sich nicht mit den Terminen der Geschwister überschneiden. In seinem Beispiel hebt Tim Berners-Lee besonders die Rolle des sogenannten Software Agenten hervor. Dieser Agent wird von beiden Geschwistern beauftragt, im Semantic Web die relevanten Informationen zu finden. Laut Tim Berners-Lee nimmt der Software Agent eine tragende Rolle im Semantic Web ein. Die Potenz des Semantic Web wird erst entfaltet, wenn viele Entwickler sich auf die Entwicklung eines Programms konzentrieren, die einzig und allein die Aufgabe hat, Web Inhalte aus verschiedenen Quellen zu sammeln, weiterzuverarbeiten und die Informationen mit anderen Programmen auszutauschen. Je mehr solcher Programme im Semantic Web kursieren, desto effektiver arbeitet ein Software Agent31. Die beiden Autoren Grigoris Antoniou and Frank van Harmelen griffen in ihrem Buch „A Semantic Web Primer“32 das Beispiel von Tim Berners-Lee auf und erläuterten die Problematik der Arztsuche in der praktischen Umsetzung. Eine gewöhnliche Webseite wird mittels der Auszeichnungssprache33 HTML aufgebaut und strukturiert. Solche Hypertext Dokumente werden dann von einem Webbrowser in terpretiert und dargestellt. Um noch einmal auf das Arztbeispiel zurückzugreifen, zeigt folgender HTML Code grob und beispielhaft die Einstiegsseite einer Arztpraxis:

Für Menschen ist diese Information lesbar und konsumierbar. Der Inhalt besteht aus einem Haupttitel, einem Text, einem Untertitel und Informationen zu den möglichen Terminen. Maschinen wären nicht fähig oder nicht zuverlässig genug, diese Angaben zu interpretieren und zu verarbeiten. Die stichwortbasierte Suche kann zwar die Schlüsselwörter wie „Physiotherapeut“ oder „Sprechzeiten“ herausfiltern, jedoch würde die stichwortbasierte Suche den Unterschied zwischen der Sekretärin und dem Therapeuten nicht erkennen. Im Semantic Web werden diese Informationen zunächst in einer speziellen Struktur zur Verfügung gestellt, mit der die Maschinen auf verlässliche Weise arbeiten können. Wandelt man das vorherige Beispiel in eine spezielle Struktur um, könnte die Lösung wie folgt aussehen:

Wie aus dem Beispiel zu entnehmen ist, sind wichtige Informationen wie der Name des Therapeuten oder der Sekretärin mit Hilfe von XML ausgezeichnet, damit die Maschinen den Unterschied erkennen können. Solche Daten bezeichnet man auch als Metadaten, also Daten, die Aussagen über andere Daten machen. Diese Struktur ist für Maschinen erkennbar, sie können damit auf zuverlässiger Art und Weise weiterarbeiten. In Kapitel 3.1 wird näher auf XML eingegangen. Da im Web verschiedene Webseiten ihre Informationen unterschiedlich aufbereiten, müssen einheitliche und offene Standards definiert werden, um eine Interoperabilität zwischen den verschiedenen Plattformen und Anwendungen zu erreichen. D. h. es muss eine Möglichkeit gewährleistet werden, die Informationen aus verschiedenen Quellen über Anwendungs- und Datenformatgrenzen hinweg zusammenzutragen, auszutauschen und ggf. miteinander in Verbindung zu setzen. Um genau solche Standards für das Web zu definieren, rief das W3C das Semantic Web Projekt ins Leben. Aus dem Projekt entstanden die grundlegenden Standards für Informationsspezifikationen wie XML, RDF(S)34 und OWL. Die XML Technologie ist zwar nicht dem Semantic Web sondern eher dem herkömmlichen Web zuzuordnen, jedoch bietet XML eine syntaktische Grundlage für das Semantic Web. Im Gegensatz dazu sind RDF(S) und OWL sogenannte Ontologie-Sprachen, die speziell für das Semantic Web entwickelt wurden. Dem Begriff Ontologie begegnet man sehr oft im Semantic Web. In Zusammenhang mit dem Semantic Web bedeutet Ontologie Wissensbasis, die nichts anderes ist als ein in RDF(S) und OWL erstelltes Dokument, welches Wissen repräsentiert. In Kapitel 2.4 wird auf den Begriff Ontologie eingegangen. Ein weiteres Merkmal im Semantic Web ist die automatische Schlussfolgerung neuer Fakten aus vorangegangenen Informationen. Die englische Bezeichnung lautet „Reasoner“ – es geht darum, implizites Wissen in explizites Wissen umzuwandeln35. Dies behebt das Problem des impliziten Wissens, das man sehr oft in der Anwendung des herkömmlichen Web findet. Diese Thematik stammt aus dem Bereich der formalen Logik. Mit der Idee des Semantic Web traten auch einige Missverständnisse auf. Es gibt zahlreiche Äusserungen36, die behaupten, durch das Semantic Web können Maschinen bzw. Computer tatsächlich in die Lage versetzt werden, die Bedeutung von Informationen zu verstehen. Die Grundidee des Semantic Web verfolgt jedoch eine moderate und einfache Zielsetzung: Mit Hilfe des Semantic Web sollen die Probleme des herkömmlichen Web behoben werden, ohne das vorhandene Web zu ersetzen oder eine parallele Technologie 34s. S. 30 K. 3.1.1 35[AH08], S. 13 36[HKRS08], S. 12 Diplomarbeit Egy Rizky Utama Ridwan 2 Semantic Web 24 neben dem klassischen Web zu etablieren. Der Grundgedanke besteht darin, Wege und Methoden zu finden, Informationenen in eine spezielle Form zu bringen und zu repräsentieren, mit der Maschinen auf eine zuverlässige Art und Weise umgehen können. Im Semantic Web geht es nicht nur um das Verstehen der Informationen seitens der Maschinen, die Maschinen sollen vielmehr die Informationen so aufbereiten, dass die Menschen einen Nutzen daraus ziehen können. Ziel ist eine bessere Kommunikation bzw. Interaktion zwischen Mensch und Maschine. Aus der Sicht von Tim Berners-Lee kann mit der Unterstützung von Ontologien das Arbeiten im Web verbessert und aufgewertet werden. Ontologien können benutzt werden, um die Genauigkeit und Treffsicherheit bei einer Websuche zu erhöhen.