

Komplexität massiv digitaler Datenbestände und deren Konsequenz

– 4. LEONARDO Sommerschule zum Thema Komplexität

Prof. Dr. Christoph Schommer

5. – 9. Juli 2010

Inhalt

Im Zeitalter des Internets, der weltweiten Kommunikation und der immer leistungsfähigeren Computer können sich digitale Daten kontinuierlich und ständig ansammeln. Die Daten werden gespeichert, korrigiert, gepflegt, fusioniert, ver- und gekauft, erhoben, visualisiert, etc., und in immer größerem Maße für analytische Zwecke genutzt. Im Workshop beschäftigen wir uns zunächst generell mit der Existenz und der Komplexität von Daten und den damit verbundenen Konsequenzen: Wie bestimmen komplexe Datenbestände unser Leben? Welche Vorteile können wir gewinnen oder werden wir zum 'gläsernen Kunden'?

Organisation

Der Workshop findet in deutscher Sprache statt. Die folgende Tabelle zeigt den zeitlichen Ablauf des Workshops:

- **Montag, 5. Juli 2010:** Am Vormittag wird das Thema des Workshops vorgestellt und die Arbeitsgruppe gebildet. Der Nachmittag beginnt mit einer grundlegenden Diskussion um das Thema und mündet in eine Reihe von Fragen, die als Grundlage für die weitere Diskussion dienen. Wir treffen uns um 14h00 an dem unten angegebenen Treffpunkt.
- **Dienstag, 6. Juli 2010:** An diesem Tag findet kein Treffen statt.
- **Mittwoch, 7. Juli 2010:** Der Nachmittag steht zum selbständigen Arbeiten zur Verfügung. Es findet kein Treffen an diesem Tage statt.
- **Donnerstag, 8. Juli 2010:** Der Nachmittag steht zum selbständigen Arbeiten zur Verfügung. Wir treffen uns um 14h00 an dem unten angegebenen Treffpunkt.
- **Freitag, 9. Juli 2010:** Am Vormittag hat die Gruppe Gelegenheit, die Präsentation vorzubereiten. Diese findet im Laufe des Nachmittags statt.

Leitfragen

- Was sind Daten? Gibt es eine einheitliche Definition dazu? Was ist der Unterschied zu Information und zu Wissen?

- Was sind Gründe für die steigende Komplexität von Daten? Welche Möglichkeiten gibt es, Daten zu sammeln? Welche Arten von Daten existieren überhaupt? Trägt eine Fusion von Daten zu dieser Komplexität bei?
- Welche **Gefahren/Risiken** sehen Sie im Hinblick auf den Umgang mit Daten? Beispiel: **soziale Netzwerke** wie Facebook, Twitter, etc.: ist hier der Schutz der Privatsphäre ein ernst zu nehmendes Anliegen? Kennen Sie das Datenschutzgesetz Luxemburg's?
- Welche **Chancen/Probleme** sehen Sie im Hinblick auf den Umgang mit Daten, zB. im Zusammenhang mit dem Thema Sicherheit, Betrugserkennung, und Verbrechensbekämpfung oder im Einsatz einer mobilen Patientenakte? Warum sind Sie hier für einen Datenaustausch?
- Um die Komplexität von Daten zu „beherrschen“, bedient man sich des Einsatzes analytischer Methoden und Verfahren. Ein Beispiel hierfür ist das **Data Mining**. Was versteht man darunter?
- Folgende Stichwörter treten häufig im Zusammenhang mit Daten auf: **Datenqualität – Maschinelles Lernen – Business Intelligence** sowie diverse Anwendungen wie **Kundenprofilierung – Customer Lifetime Value – Warenkorbanalyse** – etc. Was versteht man darunter?
- In Bezug auf Datenkomplexität und deren Auswertung: wie stehen Sie zu spezielle Karten, wie zB. die von „Auchan“ oder „IKEA“ angebotenen Punktesammelkarten? Welche Konsequenzen ergeben sich daraus?

Ziel des Workshops

Gegenstand des Workshops ist eine Untersuchung des Themas „Komplexität von Daten und deren Konsequenz“ aufgrund der oben gegebenen Leitfragen. Ziel ist ein Dokument (Präsentation), das die Datenkomplexität motiviert, die Chancen und Risiken im Umgang mit Daten beleuchtet, sowie potenzielle Anwendungsgebiete im Sinne des Data Mining adressiert. Als potenzielle Maßnahmen können neben der Recherche von Dokumenten auch Analysen durchgeführt werden.

Treffpunkt

Genereller Treffpunkt: Gebäude G, 1. Stock, linker Flügel. Je nach Wetterlage finden die gemeinsamen Treffs entweder drinnen oder draußen statt.

Literatur

- [1] Nationale Kommission für den Datenschutz in Luxembourg: siehe Webseite (stand Juni 2010) unter <http://www.cnpd.public.lu/de/index.html>
- [2] C. M. Andersen, S. Bayerl, G. Bent, J. Lee, C. Schommer: Mining your own Business. IBM Redbook Series. Vol. 1 (Retail), Vol. 2 (Finance), Vol. 3 (Telecommunication), Vol. 4 (health Care).
- [3] C. Schommer: Anwendung von Data Mining Shaker Verlag. 2003. 152pp.