

Geleitwort

Fallbasiertes Schließen, englisch Case-Based Reasoning (CBR), wurde in den 80er Jahren als ein Ansatz der Künstlichen Intelligenz entwickelt. Heute stellt es eine etablierte, wenn auch noch nicht so weit verbreitete Technik dar, um Wissen vieler Domänen über Fälle zu repräsentieren und zur Verfügung zu stellen. Der einfache Grundgedanke des CBR besteht darin, dass Lösungen auftretender Probleme am einfachsten und schnellsten durch Übernahme und eventuell Adaption der Lösungen ähnlicher Probleme, die in der Vergangenheit gelöst wurden, gefunden werden können.

Die frühen Anwendungen des CBR können als Alternativen zu regelbasierten Expertensystemen charakterisiert werden. Den letzteren gegenüber weisen sie den Vorteil der leichteren Wissensakquisition und Systemwartung auf. Die Fälle werden durch vorgegebene Attribute beschrieben, denen Werte zugeordnet werden. Die Attribute stehen oft in hierarchischer Beziehung zueinander. Dieser Ansatz wird heute genauer als Strukturelles CBR bezeichnet. Da ein erheblicher Teil des expliziten Wissens in Texten vorliegt, wurde in den letzten Jahren, insbesondere mit dem gewachsenen Interesse an einem aktiven Wissensmanagement in Organisationen, CBR auch auf unstrukturierte Fälle, in Form von Texten, angewendet. Dabei steht das Wiederauffinden von Texten im Vordergrund des Interesses, während die Adaption der vorgeschlagenen Lösung hier keine Rolle mehr spielt. Man spricht in diesem Fall von Textbasiertem CBR, und es kommen auch Methoden aus der Informationswiedergewinnung, englisch Information Retrieval, zum Einsatz.

Im vorliegenden Buch entwickelt der Autor einen Ansatz, der als eine Mischung aus Strukturellem und Textbasiertem CBR beschrieben werden kann. Die „Fälle“ liegen in Form von Texten vor. Die Beziehungen unter ihnen werden jedoch nicht etwa durch automatisches Indizieren implizit hergestellt, sondern die Texte werden einer vom Autor

entwickelten Taxonomie zugeordnet. Die Taxonomie bildet einen Teil des Wissens einer Domäne ab. Diese Taxonomie besteht aus mehreren Begriffshierarchien, je einer pro Attribut. Die in einem Text vorkommenden Attributwerte werden mit Hilfe gängiger automatischer Verfahren der Verarbeitung natürlicher Sprachen ermittelt. Die Texte selbst werden in der so geschaffenen Fallbasis nicht abgespeichert, sondern nur ihre Zuordnungen zu den Begriffshierarchien. Die Abspeicherung erfolgt in einem relationalen Datenbankmodell, das einen Großteil der Semantik der Taxonomie abbildet. Im Gegensatz zu manchen früheren Vorschlägen erlaubt der hier entwickelte Ansatz auch eine mehrfache Zuordnung eines Textes zu einem Attribut bzw. zu einer Begriffshierarchie.

Der Vorteil einer taxonomiebasierten Suche gegenüber einer nur stichwortbasierten Suche liegt darin, dass Texte gefunden werden können, die Ober- oder Unterbegriffe der gesuchten Stichworte enthalten, wobei sich je nach Wunsch diese Generalisierung bzw. Spezialisierung über mehrere Ebenen erstrecken kann. Die in der Taxonomie abgespeicherten Ähnlichkeiten der Begriffe auf der gleichen oder auf verschiedenen Ebenen stellen einen Teil des Domänenwissens dar. Die Begriffshierarchien können auch sehr gut für die Ausweitung (Relaxation) von Anfragen genutzt werden, die zu keinen oder zu wenigen Treffern führen, wie es der Autor zeigt.

Der theoretisch entwickelte Ansatz wurde auch implementiert und zur Demonstration auf Wirtschaftsmeldungen angewandt. Die Überlegenheit des Ansatzes gegenüber einer nur stichwortbasierten Suche wird an diesem Beispiel eindrucksvoll vorgeführt.

Paul Alpar