Natürlichsprachlicher Entwurf von Informationssystemen

Erich Ortner
Universität Konstanz, Informationswissenschaft
Lehrstuhl für Informationsmanagement
Postfach 5660
78434 Konstanz
Tel.: 07531-88-2976
Fax: 07531-88-2601
e-mail: ortner@inf-wiss.uni-konstanz.de

In dem Beitrag werden verschiedene linguistische Methoden der Informationssystementwicklung vorgestellt. Dabei wird der Versuch unternommen, sie nach einem einheitlichen Schema (Tabelle 1) zu klassifizieren. Zum Schluß wird auf einen Workshop zum "Natürlichsprachlichen Entwurf von Informationssystemen" hingewiesen, der vom 28.05. - 31.05.1996 in Tutzing stattfinden wird.

1. Einleitung

wie beispielsweise bei Diagrammmethoden aus "Knoten" und "Kanten", sondern aus "Begriffen" (Fach- oder Themenwörtern), die zu "Aussagen" (Sätzen) verbunden werden, entwickelt wird. Die ersten Entwicklungsergebnisse sind Aussagensammlungen, aus denen alle weiteren Ergebnisse einer Informationssystemlösung - z. T. automatisiert - abgeleitet werden können.

Diese Entwicklung setzte Anfang der 80er Jahre ein, weil mit Konzepten wie "rapid prototyping", dem Einsatz "graphischer Entwurfsmethoden" oder einem unter "demokratischen Vorzeichen" praktizierten (benutzer-)partizipativen Entwicklungstil entscheidende Qualitätsverbesserungen bei der Erstellung von Informationssystemen nicht mehr erzielt werden konnten. Vor allem die fachliche Software-Qualität (z. B. Konsistenz der Begriffsbildung bei Benutzern und den implementierten Systemen) läßt sich mit diesen Ansätzen nicht wesentlich verbessern, und der ingenieurmäßige Arbeitsprozeß - methodisches, auf Ideen und verantwortbaren Entscheidungen basiertes, rationelles Problemlösen - wird bei manchen Ansätzen teilweise entstellt dargestellt.


48
2. Arbeiten zu linguistischen Methoden der Informations-
systementwicklung

Die verschiedenen linguistischen Ansätze in Tabelle 1 können zunächst durch
die Beantwortung der Frage eingeteilt werden, ob eine existierende
(natürliche) Sprache zur Entwicklung der Informationssysteme eingesetzt
wird, oder ob eine "quasi-natürliche" Sprache zu diesem Zweck gemeinsam mit
Benutzern neu aufgebaut wird. Linguistische Entwicklungsmethoden vom
ersten Typ werden empirische Ansätze und linguistische Methoden vom
zweiten Typ konstruktive Ansätze genannt. Bei den empirischen Ansätzen
können empirisch-analytische und empirisch-experimentelle Methoden unter-
schieden werden.

Empirisch-analytische Methoden stellen im Entwicklungsprozeß natürlichsprachliche Aussagen der Benutzer über Sachverhalte eines Anwendungsge-
biets mit künstlichen formalen Mitteln (z. B. mit mathematischen Sprachen,
Diagrammsprachen oder einer Logiksprache) "exakt" dar. In empirisch-expe-
rimentellen Projekten werden Kommunikationsverhältnisse, wie sie sich z. B.
im Bürobereich ergeben, mittels protokoliert Sprechakte simuliert. Diese
Ansätze wurden auf Basis von Sprechakttheorien - z.B. in Anlehnung an

Bei den empirisch-analytischen Ansätzen werden die entwickelten Systeme
aus der Intension der (Benutzer-)Fachbegriffe "interpretativ" erarbeitet. Bei
empirisch-experimentellen Ansätzen kommt man primär über die Extension
der (Benutzer-)Fachbegriffe zu einem Systementwurf.

Als Intension bezeichnet man seit Carnap [9] die Bedeutung oder auch die
Merkmale bzw. die Definition eines Begriffs. In der Datenmodellierung wer-
den z. B. Attribute, die zu einem Objekttyp (Fachbegriff) zusammengefaßt
werden, die Intension eines Objekttyps [45] genannt. Unter der Extension
eines Begriffs faßt man alle Objekte, für die ein Begriff zutrifft (die unter den
Begriff fallen), zusammen. In der Datenmodellierung werden die Ausprä-
gungen eines Objekttyps - also die Beschreibungen existierender (realer) Ob-
jeke (Tupel, Datensätze) - seine Extension [45] genannt.

Konstruktive linguistische Methoden (Tabelle 1) sind nicht empirisch (im Sin-
ne von "aufdecken"), sondern normativ (im Sinne von "herstellen") definiert.

49
<table>
<thead>
<tr>
<th>Position</th>
<th>empirisch</th>
<th>analytisch</th>
<th>experimentell</th>
<th>konstruktiv</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>- Colombetti et al. '85 [12]</td>
<td></td>
<td>- Orther '82 [35]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Bulte/la/ van de Riet '92 [8]</td>
<td></td>
<td>- Orther/Stöliner '89 [36]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Métais et al. '83 [31]</td>
<td></td>
<td>- Vogler '94 [44]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Tjoa/Berger '83 [43]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wissens-repräsentation</td>
<td>- Weigand '90 [48]</td>
<td>- Gunia '94 [18]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Dignum/Van de Riet '91 [13]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Gerst '92 [17]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Saeki et al. '86 [36]</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Kristan '94 [24]</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Mich/Garfignano '95 [32]</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Büroinformationssysteme (Groupware, Workflow-Systeme)</td>
<td>- Lehtinen/Lyytinen '86 [27]</td>
<td>- Winograd '87/88 [50]</td>
<td>- Müller-Merbach '83 [33]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Flore et al. '88 [16]</td>
<td>- Orther '94 [37]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Auramäki et al. '88 [4]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Yonezaki '89 [52]</td>
<td>- Holmgård/Andersen '91 [31]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Roland/Proix '92 [38]</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Lachy/Janson '94 [25]</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tabelle 1: Arbeiten zu linguistischen Methoden der Informationssystementwicklung
Bei konstruktiven Ansätzen wird eine reglementierte Sprache zur Informationssystementwicklung zusätzlich hergestellt, während bei empirischen Ansätzen eine vorhandene Sprache im Entwicklungsprozeß (mit formalen Mitteln) "aufgedeckt" wird. Bei konstruktiven Ansätzen besteht ein Ziel oft darin, zwecks besserer Entwicklung und Nutzung der Informationssysteme, die fak-tische Sprachgenese in den Anwendungsbereichen um eine rekonstruierte normative Genese der Benutzerfachsprachen zu ergänzen.

Bezüglich einer Einteilung der Arbeiten zum natürlichsprachlichen Entwurf von Informationssystemen in Tabelle 1 kann man feststellen, daß empirische Ansätze im allgemeinen eine formale Sprache (Grammatik) und konstruktive Ansätze eine materiale Sprache (Grammatik und Fachlexikon) zur Entwicklung von Informationssystemen einsetzen [37].

2.1. Empirischn-analytische Methoden der Informationssystem-entwicklung


Bei konstruktiven Entwicklungsmethoden werden im Vergleich dazu öfter materiale Entwicklungssprachen eingesetzt, die aus einer Grammatik (formaler Teil) und aus einem mit rekonstruierten Fachbegriffen "gefüllten" Lexikon (materieller Teil) bestehen. Eine Unterscheidung zwischen Objekt- und Metasprachen ist bei ihrem Einsatz nicht erforderlich.

In der Informatik als Ingenieurdisciplin überwiegen natürlich die formal-sprachlichen Ansätze, während in traditionellen Ingenieurdisciplinen - z. B. im Maschinenbau - materiale Sprachen zum Einsatz kommen. Hier werden "darstellende Geometrie" und "technisches Zeichnen" als formaler Teil
(Grammatik) einer Konstruktionssprache und die Teildisziplin "Maschinen-
elemente" als ihr materielter Teil (Lexikon) angesehen.

Einem formalsprachlichen Ansatz folgend schlägt z. B. Chen [10] zur besseren
Verständigung zwischen Entwicklern und Benutzern vor, den Informationse-
darst einer Anwendung zunächst in Englisch zu dokumentieren. Die englischen
Beschreibungen werden dann von Datenbankdesignern in Entity-Rela-
tionship-Diagramme (oder ähnliche formale Repräsentationsansprachen) "um-
gewandelt". In diesem Beitrag [10] untersucht Chen die (syntaktische) Über-
einstimmung zwischen englischen Satzstrukturen und Entity-Relationship-
Diagrammen und schlägt elf Regeln für die 'Übersetzung' vor. Die Regeln für
die Ermittlung (Rekonstruktion) der Fachbegriffe bleiben dabei allerdings im
Dunkeln. Eine fachbegrifflich unreflektierte natürlichsprachliche Repräsen-
tation wird in eine fachbegrifflich unreflektierte diagrammsprachliche Re-
präsentation transformiert. Die grundlegenden Konstrukte des Englischen
- Substantiv, Verb, Adjektiv, Adverb, Gerundium, Satz - finden Gegenstücke in
der Entity-Relationship-Diagrammtechnik.

Beim Entwurf von Informationssystemen werden Wörter (Fachtermini) mit
einer festen Bedeutung benutzt. Dazu zählen Schlüsselwörter und Wörter,
welche vom Entwerfer zur Bezeichnung von Entitäten, Beziehungen, Attribu-
ten, Werten, Akteuren etc. ausgewählt wurden. Gewöhnlich haben diese
Wörter eine bestimmte Bedeutung, und manchmal stimmt ihr Gebrauch in
den Anwendungsbereichen mit dieser Bedeutung überein. In einem Beitrag
von Buitelaar und van de Riet [8] wird beschrieben, wie ein Lexikon, welches
diese Wörter und ihre Bedeutung enthält, zusammen mit einem CASE
(Computer Aided Software Engineering) - Tool erfolgreich eingesetzt werden
cann.

Ein Werkzeug zum konzeptionellen Entwurf wird in [43] vorgestellt, das Be-
darfspezifikationen in natürlicher Sprache in Konstrukte des erweiterten
Entity-Relationship-Approach transformiert. Ein Parsing-Algorithmus in
Verbindung mit einer Grammatik und einem Lexikon wird entworfen, um den
Anforderungen an das Werkzeug zu genügen. Die Ergebnisse des Parangs
werden anhand von Regeln und Heuristiken weiter verarbeitet, womit eine
"Beziehung" zwischen linguistischem Wissen und Entwurfsgebrihnissen aufge-
baut wird. Das Werkzeug arbeitet interaktiv, um mehrdeutige, unvollständige
und redundante Informationen handhaben zu können.

52


Die Fähigkeit von Computer-Anwendungen, Informationen zu verarbeiten und mit Benutzern zu kommunizieren sind stark eingeschränkt durch ihre
Unfähigkeit, "freiformatig" geschriebene oder gesprochene Sprache zu analysieren oder zu erzeugen. Während die Techniken, welche sich mit sehr be-schränkten "formalen" Sprachen befassen, gut verstanden werden, sind Techniken, welche sich mit geschriebenem Text oder gesprochem Wort befassen, noch nicht ausgereift. Die Verarbeitung solcher Informationen wird gemeinhin als "Natural Language Processing" (NLP) bezeichnet. Das LOLITA-System (large-scale object-based linguistic interactor and analyser) ist ein prototypisches NLP-System, das geschriebenes Englisch analysieren und erzeugen kann [32]. Einerseits dient LOLITA beim Informationssystementwurf der Vorverarbeitung, um Text zu korrigieren, zu selektieren und zu normalisieren. Andererseits unterstützt LOLITA die "automatische "Konstruktion objektorientierter Modelle anhand der natürlichsprachlichen Bedarfspezifi- kation, repräsentiert in Form semantischer Netze [26].


Ein Ansatz von Rolland und Proix [38] basiert auf der Prämisse, daß Wissensakquisition stark mit der Manipulation von Sprache zusammenhängt. Er stellt einen Versuch dar, die Sammlung von Aussagen zur Spezifikation einer Problemstellung, ihre Interpretation und ihre Modellierung durch den Gebrauch einer linguistischen Methode zu verbessern. Es wird vorgeschlagen, die Problemspezifikation für die Informationssystementwicklung durch Sätze in natürlicher Sprache auszudrücken. Es wird gezeigt, wie mit einem linguistischen Ansatz, der auf dem CASE-Gedanken beruht, automatisch die Modellierung eines Informationssystems durchgeführt werden kann. Der linguistische Ansatz und seine Implementierung als Expertensystem - bekannt als OICSI (französisches Akronym für "intelligentes Werkzeug für die Informationssy-
stementwicklung") - werden im Detail erläutert. Die zentrale These lautet, daß der linguistische Ansatz allgemeingültig ist, und zwar in zweifacher Hinsicht. So kann er an verschiedene Modellierungstechniken angepaßt und zusätzlich für ein weites Problemspektrum genutzt werden.

2.2. Empirisch-experimentelle Methoden der Informationssystementwicklung


Im anderen Falle (nur Sprechhandlungen liegen vor) wird von Entwicklern beim Systementwurf die Bedeutung relevanter Fachtermini in Gesprächen mit Benutzern über Handlungen und ihre Gegenstände in Anwendungsbereichen ermittelt. Hier ist die "natürlich" Kontrolle der Aussagen (Sprechhandlungen) durch das begleitende konkrete Handeln nicht mehr möglich, und es ist nicht ausgeschlossen, daß sich Mißverständnisse ergeben.


55
Erhebungsverfahren wie die Simulation von Arbeitssituationen in Anwendungsbereichen - mit sprachlich begleitender Protokollierung der Sachverhalte durch ausführende Personen oder Beobachter - eignen sich zur empirischen Analyse eines Fachentwurfs. Dagegen kann z. B. durch Interviews, Fragebögen oder "Beispieldiskussionen" (auf rein sprachlicher Ebene) ein Fachentwurf epipraktisch überprüft werden.


Die Entwicklung eines (Büro-)Informationssystems geschieht gewöhnlich aus einer ganz bestimmten Perspektive, die die Entwürfe, welche gestellt werden, sowie die Lösungen, welche gesucht werden, von vornherein beeinflußt. In [50] führt Winograd Perspektiven ein, welche auf der Interpretation der Sprache als Handlung beruht. Die Konsequenzen für den Systement-


2.3. Konstruktive Methoden der Informationssystementwicklung

Konstruktive Methoden der Informationssystementwicklung haben den Aufbau einer - zu der vorhandenen Fachsprache eines Anwendungsgebiets kom-


In (36) stellen die Autoren ein in einem Software-Unternehmen auf obigem Ansatz (46), (35) basierendes implementiertes Verfahren zur Datenmodellierung (Objekttypenmethode) vor, bei dem der methodische Entwicklungsprozeß mit der Sammlung relevanter (natürlichsprachlicher) Aussagen über Sachzusammenhänge in den Anwendungsbereichen beginnt. Es folgt die Klärung und Rekonstruktion wichtiger Fachbegriffe, wobei Regeln zur Behandlung von Synonymen, Homonymen, Äquivalenten, Vagheit und falscher Bezeichner (Begriffsverwirrung) angegeben werden. Aus der sich auf diese Weise ergebenden Aussagensammlung können durch Analyse der Satzstrukturen und unter Be-


Die Entwicklung von Anwendungssystemen erfordert eine systematische Vorgehensweise bereits in den frühen Projektierungsphasen. Eine grundlegende Schwäche der meisten Modellierungsmethoden für den Fachentwurf ist die mangelnde benutzerneffekte systematische Konstruktion der Entwicklungsergebnisse aus dem Fachwissen der Anwendungsbereiche heraus. In [40] wird ein Ansatz vorgestellt, der die objektorientierte Spezifikation eines Fachkonzeptes auf der Grundlage einer Rekonstruktion der Fachbegriffe eines Anwendungsbereichs in Zusammenarbeit mit den Anwendern ermöglicht. Unterteilt in die Aufgabenfelder Spezifikation, Repräsentation und Integration werden Lösungskonzepte für diesen "terminologie-basierten Ansatz" entwickelt und in
ein Vorgehensmodell für den Fachentwurf mit den Phasen "Rekonstruktion" und "Spezifikation" integriert.


3. Workshop zum natürlichsprachlichen Entwurf von Informationssystemen

Ein methodisch-natürlichsprachlicher Beginn der Entwicklung von Informationssystemen wird in der Software-Industrie [36] sowie bei Anwendern in manchen Unternehmen [6], [23] bereits praktiziert. Daneben gibt es - wie Tabelle 1 zeigt - zahlreiche Projekte, die eine Entwicklung von Methoden und Modellen sowie die Bereitstellung von Werkzeugen zur Unterstützung des natürlichsprachlichen Entwurfs von Informationssystemen zum Ziele haben. Die GI-Fachgruppe 2.5.2 EMISA (Entwicklungsmethoden für Informationssy-

Mit dem Workshop ist geplant, eine Diskussion innerhalb und zwischen Teil-disziplinen wie Requirements Engineering, Information Engineering, Datenbank-Anwendungsentwicklung, Wissensrepräsentation, (Computer-)Linguistik und Wissenschaftstheorie anzuregen. Als Themekomplexe, die zur Orientierung für Beiträgeinreichungen dienen sollen, werden

- Grundlagen methodisch-natürlichsprachlicher Systementwicklung,
- Entwicklung einzelner Methoden und ihre Integration in Vorgehensmodelle,
- Werkzeugentwicklungen und Werkzeugeinsatz,
- Projekte bei Anwendern sowie in Software-Unternehmen


Literatur


63


64


