T. Gross (Hrsg.): Mensch & Computer 2007: Konferenz für interaktive und kooperative Medien. München: Oldenbourg Verlag. 2007, S. 233

Visualisieren und Editieren komplexer Ontologien mit einer adaptiven Baumkomponente

Peter Hüsken, Jürgen Ziegler

Universität Duisburg-Essen

Zusammenfassung

In diesem Paper wird eine Konzeptstudie einer Baumkomponente für die Visualiserung von Ontologie-Strukturen vorgestellt. Neue grundlegende interaktive Techniken des hier eingeführten SmartTrees werden im Hinblick auf eine benutzerfreundliche Ontologie-Exploration beschrieben. Darüber hinaus werden Grundlagen für die Realisierung einer adaptiven Variante diskutiert. Verschiedene Größenskalierungen für die Darstellung von Konzepten unterstützen den Nutzer bei der Unterscheidung wichtiger oder weniger wichtiger Konzepte.

1 Einleitung

Die Komplexität der Ontologieentwicklung liegt in der typischerweise hohen Anzahl an Konzepten bzw. Instanzen – umfangreiche Ontologien wie etwa GeneOntology (Smith et al. 2003) umfassen mehrere zehntausend Konzepte – und der komplexen Vernetzungsstruktur begründet. Als reine Netzwerkstruktur (allgemeine gerichtete Graphen) ergeben umfangreiche Ontologien unüberschaubare visuelle Graphstrukturen, die als Ganzes nicht visualisierbar sind. Existierende Ontologie-Editoren wie Protege (Grosso et al. 1999) stützen sich deshalb zumeist auf baumstrukturierte, hierarchische Teilstrukturen der Ontologie ab. Eine solche Teilstruktur ist immer in Form der Klassenhierarchie der Ontologiekonzepte gegeben, die quasi das Gerüst der Ontologie bildet und an die alle weiteren Visualisierungen und weiteren Dialoge angebunden sind. Die Visualisierung und Interaktion mit solchen Baumstrukturen bildet deshalb einen der Schlüsselfaktoren für eine übersichtliche, verständnisfördernde Darstellung der komplexen Struktur und für ein effektives und effizientes Editieren. Die bislang für diese Aufgabe praktisch ausschließlich herangezogenen konventionellen Tree-Widgets (wie z. B. das aus dem Windows File Explorer bekannte) sind zur Visualisierung großer Ontologien allerdings nur bedingt geeignet, so dass wir in der Verbesserung der Prä-

sentation und Interaktion mit Tree-Widgets ein Potential zur Optimierung des Editierens komplexer Ontologien sehen.

Beim Editieren können zwei grundsätzliche Unterscheidungen gemacht werden, je nach Art des Zugriffs auf die zugrunde liegende Ontologie: Der lesende Zugriff (*Browsing*) einerseits und der schreibende Zugriff (*Editing*) andererseits. Wir beschränken uns hier auf den Browsing-Modus. Im Folgenden werden interaktive Techniken der Baumkomponente SmartTree vorgestellt, um die Aufgabe der Ontologie-Editierung benutzerfreundlicher und effizienter zu gestalten. In Kapitel 3 werden dann Grundlagen für die Möglichkeit der Benutzeradaption behandelt, so dass die Komponente sich dem angenommenen Benutzerinteresse anpassen kann.

2 Interaktive Funktionen des SmartTree

Innerhalb der Baumdarstellung repräsentieren die Knoten Konzepte der Ontologie, Konzeptinstanzen werden als gesondert gefärbte Blätter im Hierarchiebaum angezeigt. Basierend auf dem konventionellen Tree-Widget haben wir neue Funktionen für die spezifische Ontologie-Visualisierung ergänzt. Zur Gewährleistung der Übersicht auch im Falle komplexer Ontologien wird z. B. die *Fokus+Kontext*-Visualisierungstechnik (Furnas 1986) eingesetzt.

Unter Fokus+Kontext versteht man Methoden, die die Visualisierung des Datenraums derart verändert, dass der Fokus – der für den Betrachter interessante Teil der dargestellten Daten – möglichst maximiert wird, während der Kontext – die sich an den Fokusbereich anschließenden Bereiche der Visualisierung – verkleinert dargestellt werden. Der Vorteil dieser Visualisierungstechnik für den Nutzer ist die Präsentation von Detail-Informationen, während Verbindungen zu den übrigen Objekten des Kontexts angezeigt bleiben und so strukturelle Zusammenhänge (hier die vollständige Konzepthierarchie der Ontologie) für den Nutzer erkennbar sind.

Eine weitere Anpassung betrifft die gemeinsame Darstellung von Konzepten und deren Eigenschaften (engl. properties) im SmartTree. Sog. *Property-Lines* repräsentieren Konzeptrelationen mit Definitionsbereichen, in denen das aktuell selektierte Konzept enthalten ist. In der Web Ontology Language (Smith et al. 2004) werden diese zwei-stelligen Relationen auf der Konzeptmenge *Object-Properties* genannt. So kann z.B. ausdedrückt werden, dass Instanzen eines Konzepts A Instanzen eines Konzepts B verwenden. Property-Lines verknüpfen nun selektierte Konzeptdarstellung mit Konzepten im Wertebereich des Object-Properties. Referenzierte Konzepte des Kontextbereichs sollten ebenfalls hervorgehoben dargestellt werden, daher wurde die Fokus-und-Kontext-Technik erweitert: Anstelle eines Fokusbereichs, gibt es nun eine Menge von Fokusbereichen mit jeweils angrenzenden Kontextbereichen.

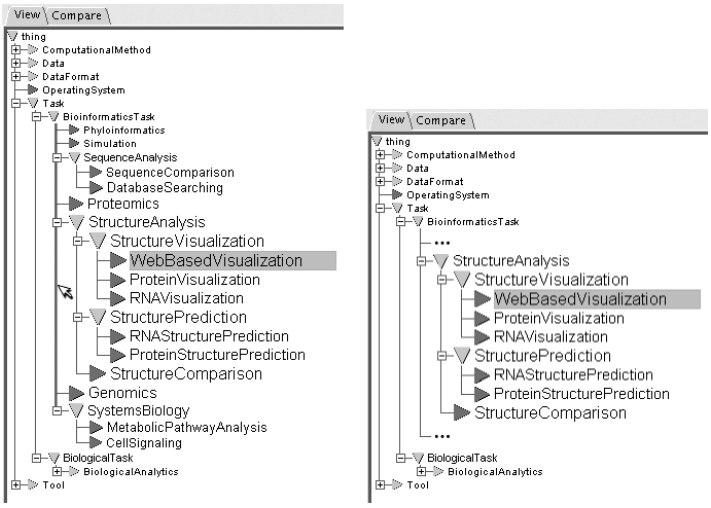

Abbildung 1: Selektion einer Ebenenlinie (links) und SmartTree nach Condense-Aktion (rechts)

Für den Fall der Weiterentwicklung einzelner Teilbäume der Konzepthierarchie können nicht relevante Teilbäume konventionell eingeklappt werden. Diese grundlegende Technik – hier *Expand & Collapse* genannt – wird in allen Tree Widgets eingesetzt. Darüber hinaus haben wir eine weitere Klapptechnik mit der Bezeichnung *Condense & Explode* für die SmartTree-Komponente implementiert: Durch Selektion der Linie, die Teilbäume mit ihrem gemeinsamen Elterknoten verbindet, werden alle Teilbäume mit gleichem direkten Elterknoten, die ausserhalb des Fokus stehen, ausgeblendet (condense) und durch ein Auslassungszeichen dargestellt. Auf der linken Seite von Abb. 1 wird die Selektion einer Verbindungslinie dargestellt, auf der rechten Seite wird der SmartTree nach erfolgter Condense-Aktion gezeigt. Die Selektion eines Auslassungszeichens führt eine explode-Aktion durch, so dass alle ausgelassenen Teilbäume dieser Stelle anschließend wieder sichtbar sind. Für den Nutzer hat dieser Klappmechanismus den Vorteil, weitere zum aktuellen Zeitpunkt unwichtigen Teile der Darstellung auszublenden, um den begrenzten Platz in der vertikalen Dimension für wichtigere Konzeptstrukturen zu verwenden. Wird dieser Klappmechanismus in Abhängikeit von den Wichtigkeitswerten der dargestellten Konzepte automatisiert (siehe Abschnitt 3.2), dann wird die Scrolling-Funktion des SmartTrees seltener benötigt und so ein effizientes Editieren unterstützt.

3 Adaptive Funktionen des SmartTree

Der bisher beschriebene SmartTree verwendet unterschiedliche Hervorhebungen von Konzeptdarstellungen auf der Grundlage der aktuellen Konzeptselektion des Nutzers und struktureller Beziehungen des selektierten Konzepts mit anderen Konzepten in der Ontologie. Ziel einer adaptiven SmartTree-Komponente ist darüber hinaus eine Abschätzung des wahrscheinlichen Nutzerinteresses und eine dementsprechende Anpassung durch Fokussierung

relevanter Konzepte. Dazu werden allen Objekten Wichtigkeitswerte (engl. *Degree-of-Interest* values (DOI)) zugeordnet. Card et al. (2002) bezeichnen die Anwendung des DOI-Konzepts für Baumdarstellungen als *logical filtering* der Knotenmenge.

3.1 Wichtigkeitswerte (DOIs) für Konzeptdarstellungen

Für die algorithmische Bestimmung der Relevanz einzelner Objekte ist die Konstruktion eines mathematischen Modells der Abhängigkeiten notwendig. Das DOI-Konzept wurde von Furnas (1986) im Zusammenhang mit sog. *Fisheye-Menu*-Visualisierungen eingeführt. Jedem Objekt der Visualisierung wird ein Wichtigkeitswert zugeordnet der sich aus zwei Teilen zusammensetzt:

1. Der *A Priori Wichtigkeit* API(x), welche die Wichtigkeit des Informationsobjektes x im Informationsraum, unabhängig vom Betrachter beschreibt. Der API-Wert ist abhängig von der Ontologiestruktur und ist daher für eine unveränderte Ontologie konstant. Veränderungen des API-Werts ergeben sich durch Editieren der Ontologie. Zwei einfache Regeln zur API-Bestimmung werden verwendet:
   - Die Wichtigkeit eines Konzeptes ist umso höher je mehr Instanzen des Konzepts existieren.
   - Die Wichtigkeit eines Konzeptes ist umso höher je häufiger es in Konzeptrelationen (Object-Properties) als Teil des Definitions- oder Wertebereichs enthalten ist. Wir vermuten, dass ein Konzept mit vielen Object-Properties häufiger instanziiert wird und so wieder eine erhöhte Wichtigkeit angenommen werden kann.
2. Der *Distanz* D(x, SP) zwischen dem Informationsobjekt x und dem aktuell selektierten Informationsobjekt (SP) des Betrachters. Eine große Distanz bedeutet, dass Konzept x weniger wichtig ist. Verschiedene Distanztypen können angewendet werden. Um den *Fisheye-Effekt* zu erzielen, welcher grundlegend für Fokus+Kontext-Techniken ist, wird eine geometrische Distanz D_{vis} verwendet. Es ist üblicherweise definiert als die Anzahl der Konzeptdarstellungen zwischen SP und x innerhalb der Baumdarstellung. Auf diese Weise erhält der Nutzer einen besseren Überblick über die benachbarten Konzepte des selektierten Konzepts. Es ist ebenfalls möglich eine taxonomische Distanz D_{tax} als ein Beispiel einer semantischen Distanz zu verwenden, die bereits Rada et al. (1989) verwendet hat. Zusammenfassend werden zwei Annahmen getroffen: (a) Konzepte sind ähnlich (geringe Distanz), wenn der kürzeste Pfad zwischen ihnen in der Taxonomie kurz ist und (b) Konzepte in höheren Ebenen der Konzepthierarchie sind genereller und daher semantisch weniger ähnlich als Konzepte tiefer liegender Ebenen. Ein kombinierter Distanztyp wird von Card et al. (2002) verwendet. Sie nehmen an, dass die Wichtigkeit eines Knotens mit der Pfadlänge zur Wurzel der Konzepthierarchie und der geometrischen Distanz zum selektierten Konzept abnimmt.

Der Wichtigkeitswert für Objekt x in Abhängigkeit vom aktuellen Fokuspunkt berechnet sich über eine Funktion F:

$$DOI(x, SP) = F(API(x), D(x, SP)) \quad (1)$$

- Gleichung 2 zeigt die allgemeine Form des F-Maßes als eine mögliche Funktion F. Das F-Maß wird im Information Retrieval zur kombinierten Bewertung der booleschen Retrievalmaße *Precision* und *Recall* verwendet. Der zusätzliche Faktor β gewichtet hier die Distanz D (β = 0: nur API(x) wird berücksichtigt, β = ∞: nur D(x,SP) ist relevant, β = 1: beide Argumente werden gleich gewichtet). Die Argumente API und D haben den Wertebereich Null bis Eins, so dass auch F_β diesen Wertebereich besitzt.

- $$F_\beta(x,y) = \frac{(\beta^2+1)\cdot x \cdot y}{\beta^2 \cdot x + y} \quad (2)$$

- Die Exploration der interaktiven Illustration ist einen iterativer Prozess, da der Nutzer auf Grundlage gewonnener Informationen seine nächsten Schritte entscheidet. Der Verlauf der Exploration wird berücksichtigt, indem Gleichung 1 folgendermaßen modifiziert wird:

$$DOI_i(x,SP_i) = \begin{cases} API(x) & : i = 0 \\ F(DOI_{i-1}(x,SP_{i-1}), D(x,SP_i)) & : sonst \end{cases} \quad (3)$$

Es sollte darauf geachtet werden schnelle Distanzberechnungen zu verwenden, da andernfalls die Performance des SmartTrees beinträchtigt wird.

3.2 Weitere Funktionen des SmartTrees

Für die Visualisierung einer Ontologie stehen nun neben der Ontologie die Wichtigkeitswerte aller Konzepte zur Verfügung. Diese werden direkt für die Größenskalierungen der Konzeptdarstellungen verwendet. Es hat sich gezeigt, dass hier eine exponentielle Skalierung im Gegensatz zur linearen Skalierung besser geeignet ist, da wenige besonders stark hervorgehobene Konzeptdarstellungen für viele Nutzer verständnisfördernder sind.

Die Verwendung von Wichtigkeitswerten ermöglicht weitere Funktionen:

Automatische Klappmechanismen für Teilbäume: Übersteigt der DOI eines Konzeptes den DOI eines Elters in der Visualisierung, dann wird der Teilbaum aufgeklappt und das Konzept angezeigt. Enthält ein Teilbaum nur untergeordnete Konzepte mit sehr viel kleineren DOI-Werten als das Wurzelkonzept des Teilbaums, dann wird der Teilbaum zusammengeklappt. Ähnlich verhält es sich mit der Condense & Explode Technik: Sobald ein Konzept den minimalen DOI-Wert besitzt, wird anstelle der Konzeptdarstellung ein Auslassungszeichen (Ellipse) angezeigt.

Häufig liegen neben einer Ontologie noch weitere externe Dokumente vor, die ebenfalls Informationen über die Domäne enthalten. Diejenigen Konzepte oder Instanzen, welche ebenfalls in einem verfügbaren Dokument beschrieben werden, werden im SmartTree durch ein gesondertes Icon dargestellt. Jedes Dokument das ein Konzept c referenziert, erhöht den DOI-Wert des Konzeptes.

4 Zusammenfassung und Ausblick

In dieser Konzeptstudie wurde das konventionelle Tree Widget hinsichtlich einer geeigneten interaktiven Ontologie-Visualisierung zur neuen Baumkomponente SmartTree weiterentwickelt, die die Konzepthierarchie einer Ontologie anzeigt. Neue interaktive Funktionen wie Condense & Explode und Visualisierungstechniken wie Property-Lines wurden vorgestellt. Um Benutzeradaption zu realisieren, wurden schwerpunktmäßig Grundlagen für die Berechung von Wichtigkeitswerten (DOI) für Konzepte von Ontologien präsentiert. Neben der Skalierung der Konzeptdarstellungen anhand der Wichtigkeitswerte wurden weitere Funktionen wie automatisierte Klappmechanismen für Teilbäume und die Unterstützung externer Dokumente beschrieben.

Literatur

Card, S. K., Nation D. (2002): Degree-of-Interest Trees – A Component of an Attention-Reactive User Interface, International Conference on Advanced Visual Interfaces.

Furnas, G. W. (1986): Generalized fisheye views. In Mantei, M. & P. Orbeto (Hrsg.), Proc. of ACM CHI Conference on Human Factors in Computing Systems, S. 16-23. New York: ACM Press.

Grosso, W. E., Eriksson, H., Fergerson, R. W., Gennari, J. H., Tu, S. W and Musen, M. A. (1999): Knowledge Modeling at the Millenium (The Design and Evolution of Protege-2000). Proceedings of the 12th Knowledge Acquisition Workshop (KAW'99)

Rada, R., Mili, H., Bicknell, E., Blettner, M. (1989): Development and application of a metric on semantic nets. IEEE Transactions on Systems, Man and Cybernetics 1989, 17-30

Smith, B.; Williams, J; Schulze-Kremer, S. (2003): The Ontology of the Gene Ontology. Proceedings of AMIA Symposium 2003.

Smith, M. K., Welty, C., McGuinness, D. L. (2004): OWL Web Ontology Language Guide. http://www.w3.org/TR/owl-guide (09.03.2007)

Kontaktinformationen

Peter Hüsken, Jürgen Ziegler
- Interaktive Systeme und Interaktionsdesign -
Universität Duisburg-Essen (Campus Duisburg), 47048 Duisburg
E-Mail: huesken | ziegler@interactivesystems.info