Probabilistic LCS in a P-Classic Implementation

Alissa Kaplunova, and Ralf Möller

1. Oktober 2007

Zusammenfassung

In this report, we describe investigations made in the context of user-adaptive information systems. We verify the idea of combining two formalisms: description logics (DLs) and Bayesian networks in order to increase the effectiveness of information retrieval. For this, we implemented the basic functionality of *P*-Classic which extends the DL Classic with probabilistic inferences. In *P*-Classic, the degree of concept subsumptions can be quantitatively expressed as a statistical value. We use this feature for the definition of the *P*LCS operator, a probabilistic "Least Common Subsumer" operator which allows for quantitative measure of concept overlap, and we show how this operator can improve the quality of information retrieval.

The software package including the prototypical implementation of *P*-Classic in Common Lisp can be obtained from `http://www.sts.tu-harburg.de/~r.f.moeller/band/pclassic.zip`.

STS Institute for Software, Technology & Systems

TUHH Hamburg University of Technology

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

1 Einleitung

Seit der Erfindung des WWW – World Wide Web –, einem Netz von sogenannten Web-Dokumenten unterschiedlicher Typen im Internet, die durch Hypertext-Verknüpfungen verbunden sind, explodierte die Anzahl der Web-Seiten in kürzester Zeit von anfänglichen 50 auf mehrere Millionen. Angesichts des rapiden Ansammelns von Daten im Internet wird der Problematik der schnellen und gezielten Recherche nach Material zu einem bestimmten Thema immer größere Bedeutung beigemessen. Die sogenannten Online-Informationsdienste spezialisieren sich dabei allein auf die Aufbereitung und Bereitstellung von Informationen im Internet. Zum Zweck des „Information Retrieval" werden im Internet zum Beispiel sogenannte Suchmaschinen eingesetzt. Eine Suchmaschine gibt dem Benutzer die Möglichkeit, durch die Angabe von Schlüsselwörtern und deren Verknüpfungen beispielsweise mittels logischer Operatoren in speziellen Online-Formularen nach den diese Schlüsselwörter enthaltenen Dokumenten zu suchen. Aus programmtechnischer Sicht verbirgt sich dahinter oft eine in einer Computersprache gestellte Anfrage an die Datenbank, in der die gesuchten Informationen gespeichert sind. Es bedarf daher u.U. guter Kenntnisse dieser Anfragesprache, um akzeptable Suchergebnisse zu erzielen. Ist eine Anfrage-Spezifikation zu exakt, werden keine Dokumente gefunden. Oftmals werden aber eher zu viele Treffer zurückgegeben, so daß die Recherche vom Gesichtspunkt der Effizienz erheblich beeinträchtigt ist.

Obwohl der Vergleich von Wörtern in Dokumenten mit Wörtern in einer Anfrage immer noch die verbreitete Form eines Suchvorgangs darstellt, sind die Ergebnisse einer solchen Suche meistens unbefriedigend und sehr stark vom Sprachvokabular des Benutzers abhängig. Ein unerfahrener Benutzer hat üblicherweise Schwierigkeiten, die gesuchte Information mit den dafür zur Verfügung stehenden Mitteln korrekt zu beschreiben.

Ein weiterer Kritikpunkt der bestehenden Informationsdienste ist unbefriedigender Umgang mit großen Datenmengen. Die Filterung von relevanten Informationen stellt eine wichtige Voraussetzung für eine effiziente und erfolgreiche Recherche dar. Darüber hinaus „weiß" das System in der Regel nichts vom Benutzer und kann entsprechend keine Vorauswahl treffen. Die Präferenzen des Benutzers werden meistens weder in dem Inhalt noch in der Präsentationsform der Ergebnisse berücksichtigt.

1.1 Motivierendes Beispiel: Benutzeradaptiver Informationsdienst

Die Mängel der traditionellen, in der Regel formularbasierten Suchdienste führten dazu, daß neue Technologien zur Informationsrecherche erforscht werden. Das Spektrum reicht hier von logikorientierten Ansätzen bis zu Fuzzy-Set basierten Anfragen und probabilistischen Techniken. Eine umfangreiche Übersicht über die wichtigsten Forschungsrichtungen in diesem Bereich findet man z.B. in [8].

Im Kontext des „intelligenten" Information Retrieval und benutzeradaptiven Präsentationsformen kann u.a. die Idee der „beispielbasierten" Suche oder „fallbasierten Informationsrecherche" untersucht und anhand eines prototypischen Web-TV-Auskunftssystems illustriert werden ([14]).

Bei der beispielorienten Suche wählt der Benutzer aus dem existierenden Datenvorrat (z.B. aus TV-Sendungen) Beispiele aus, zu denen er andere inhaltlich passende Objekte finden möchte. Anstatt komplizierte Anfragen zu formulieren, fügt der Benutzer die für ihn interessanten Objekte zu einer Art „Sammlung" hinzu. Das System versucht anhand der Beispiele einer erstellten Sammlung neue Objekte zu ermitteln, die mit der getroffenen Auswahl

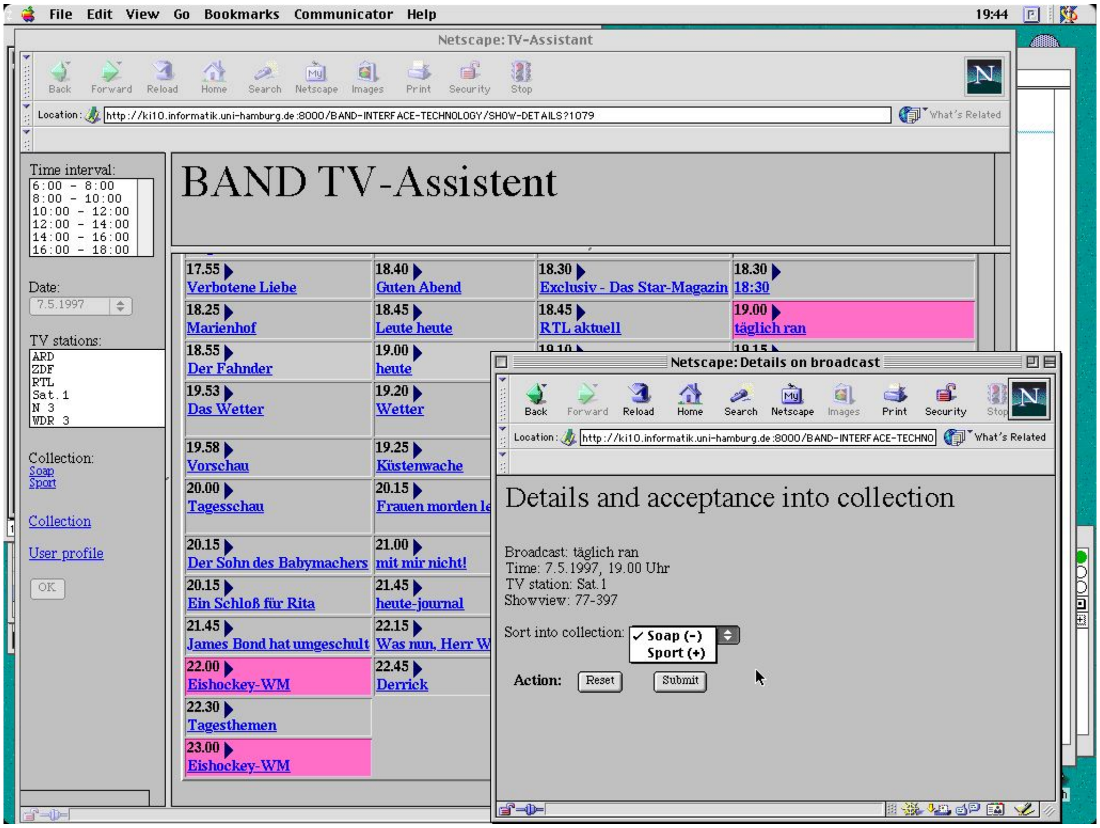

Abbildung 1: Web-TV-Assistent.

etwas Gemeinsames haben. Zur Veranschaulichung zeigt Abbildung 1 den TV-Assistenten in einem Zustand, in dem für die zusammengestellte Sammlung „Soap", die 2 Sendungen enthält - „Marienhof" und „Unter uns", weitere ähnliche Sendungen gefunden und farbig hervorgehoben werden. Die Hervorhebung ermöglicht eine perzeptive Datenfilterung. Zwar wird die Aufmerksamkeit des Benutzers durch graphische Visualisierung auf bestimmte Informationen gelenkt, aber auch andere Informationen bleiben ihm nicht vorenthalten. Durch Einfügen oder Herausnehmen von Objekten aus einer Sammlung steuert der Benutzer implizit das Verhalten des Systems, das mit neuen Vorschlägen auf die sich veränderten Präferenzen des Benutzers reagiert. Man kann vermuten, daß diese Art der Benutzeradaptivität eher Akzeptanz seitens der Anwender finden würde als die üblichen fragebogenbasierten Mechanismen zur Erstellung von Benutzerprofilen. Neben einer intuitiven Bedienung bietet ein beispielorientierter Informationsdienst mehr Transparenz, denn die Zusammenhänge zwischen den Objekten in der Sammlung und der gefundenen Informationen sind für den Benutzer eher nachvollziehbar.

1.2 Beschreibungslogiken für Benutzeradaptive Informationsdienste

Es stellt sich die Frage, mit welchen Mitteln ein Informationsdienst, der die beschriebene Funktionalität bietet, in die Praxis umgesetzt werden kann. Als erstes muß die Klarheit verschafft werden, welche Stellung dieser Dienst in einem komplexen Informationssystem annimmt. Abbildung 2 zeigt eine grobe Skizze der allgemeinen Architektur. Der Fokus unserer

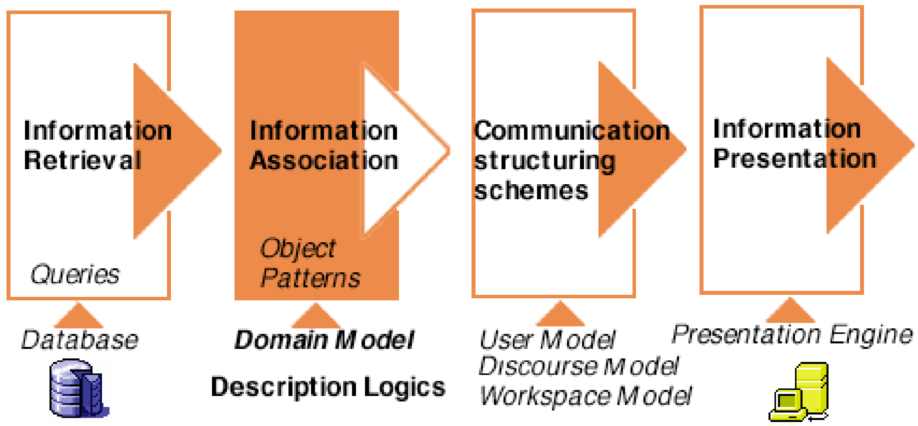

Abbildung 2: Architektur eines TV-Agenten.

Betrachtung richtet sich auf die Komponente, die als „Information Association“ bezeichnet ist. Als Basisformalismus zum Modellieren der Anwendungsdomäne wurden Beschreibungslogiken gewählt. Für den als Leitbeispiel entwickelten TV-Assistenten wurde das auf Beschreibungslogiken basierende Wissenspräsentationssystem Classic verwendet. Im Kapitel 2 wird auf die Merkmale und einige Schlüsselaspekte von Beschreibungslogiken eingegangen. An dieser Stelle sei lediglich kurz angemerkt, daß mit Mitteln einer Beschreibungslogik sich Wissen über eine Anwendungsdomäne in Form von Konzepten und Relationen repräsentieren läßt. Konzepte werden durch Klassifikation in einer Hierarchie angeordnet. Es entsteht eine sogenannte Terminologie. Bestimme Fakten und Zusammenhänge aus der realen Anwendungswelt können in Form von sogenannten assertorischen Aussagen repräsentiert werden. Ein konkretes Anwendungsobjekt (Individuum) wird als Instanz eines Konzepts oder einer Reihe von Konzepten dargestellt. In der TV-Domäne wurden z.B. Schauspieler und Filme mit den Konzepten „Darsteller“ bzw. „Film“ beschrieben. Die Beziehung zwischen Instanzen dieser Konzepte wird als Rolle „spielt-in-Film“ modelliert. Ein konkreter Film, z.B. „Goldfinger“, wurde rechnerintern als eine Instanz des Konzepts „James-Bond-Film“, ein Unterkonzept von „Film“, deklariert. Darüber hinaus implementieren an Beschreibungslogiken angelehnte Wissenspräsentationssysteme, wie das erwähnte Classic, eine Reihe deduktiver Schlußverfahren für Konzepte und Individuen.

1.3 Ergebnisse dieser Arbeit

Wenn man sich mit der Klassifizierung einer Anwendungswelt befaßt, stellt man fest, daß sich Objekte nicht immer eindeutig mit einem Oberbegriff beschreiben lassen. Ein Krimi kann z.B. ein Film wie eine Serie sein. Eine Tennis-Davis-Cup-Über-tragung würde man im allgemeinen als Individualsportart-Sendung sowie auch als Mannschaftsspiel-Sendung modellieren, da man nicht festlegen möchte, ob in der konkreten Tennis-Davis-Cup-Ausstrahlung ein Spieler gegen einen anderen Spieler oder eine Mannschaft gegen eine andere Mannschaft spielen wird. Bei einer Tennis-Davis-Cup-Sendung bestehen beide Möglichkeiten. Man würde anhand der eventuell vorhandenen heuristischen Daten behaupten, daß es eher, d.h. mit größerer Wahrscheinlichkeit ein Zweikampf wird. Generell ist es allerdings nicht eindeutig

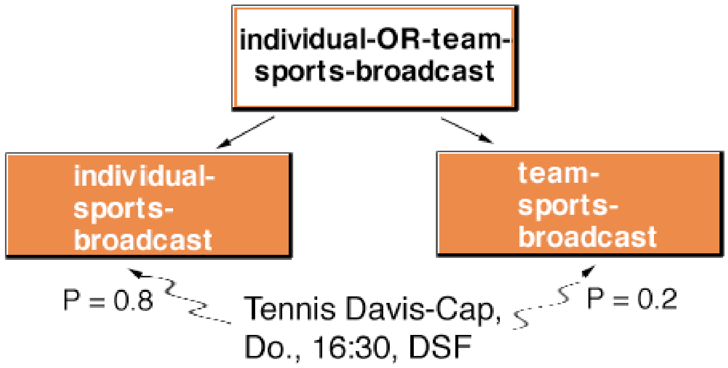

Abbildung 3: Überlappung von Konzepten.

zu beurteilen, ob über den gegebenen Sachverhalt eine oder mehrere Aussagen zutreffen oder nicht. Das Wissen ist unsicher. Andererseits gibt es durchaus Sportarten, die nur als Mannschafts- oder nur als Individualsportart bezeichnet werden können. Um eine Tensis-Davis-Cup-Übertragung dennoch klassifizieren zu können, müßte man ein neues Konzept definieren, welches das Gemeinsame von Individual- und Mannschaftssportart-Sendungen beschreibt (siehe Abbildung 3). Es scheint in der Natur des Menschen zu liegen, eher in sich überlappenden Kategorien zu denken. Eine Überlappung liegt vor, wenn mehrere Konzepte eine Instanz untereinander teilen, aber keins der Konzepte von den anderen vollständig subsumiert wird. In unserem Beispiel ist Tennis-Davis-Cup zwar eine Instanz von beiden Konzepten Individual- und Mannschaftssportart-Sendung, aber im allgemeinen ist nicht jede Instanz des Konzepts Mannschaftssportart-Sendung auch Instanz von Individualsportart-Sendung und umgekehrt. Abgesehen von psychologischen Gegebenheiten gibt es durchaus pragmatische Gründe, um die Überlappung von Konzepten in einem Wissenspräsentationssystem handhaben zu wollen. Zum Beispiel läßt sich das Wissen kompakter, d.h. mit weniger Konzepten repräsentieren [12]. Insbesondere im Kontext der benutzeradaptiven Informationsrecherche spielt die adäquate Modellierung und Verarbeitung unsicheren Wissens eine bedeutsame Rolle. Obwohl es durch Verwendung von Disjunktionen leicht möglich ist, in Beschreibungslogiken Wissen unvollständig zu beschreiben, sind Beschreibungslogiken eingeschränkt in ihrer Fähigkeit, unsicheres Wissen auszudrücken.

In dieser Arbeit wird gezeigt, daß die Kombination der Beschreibungslogik $\mathcal{ALN}$ mit einem wahrscheinlichkeitstheoretischen Verfahren unter bestimmten Voraussetzungen ein interessantes Hilfsmittel darstellt, um die Überlappung von Konzepten quantitativ auszudrücken. Dieser Ansatz wurden zum ersten Mal in [7] unter dem Namen *P*-CLASSIC, eine probabilistische Erweiterung von CLASSIC, vorgestellt. Zu den Zielen der hier vorliegenden Arbeit gehören u.a. die Untersuchung, Implementation und Evaluierung der Basisfunktionalität von *P*-CLASSIC. *P*-CLASSIC bietet ein Instrument, mit dem unsicheres Wissen über Objekteigenschaften mit Methoden der Wahrscheinlichkeitstheorie modelliert und verarbeitet werden kann. Die probabilistische Komponente von *P*-CLASSIC basiert auf dem Bayes-Netz Formalismus. Im folgenden Kapitel werden die für das Verständnis von *P*-CLASSIC relevanten Grundlagen der Wahrscheinlichkeitstheorie und Bayes-Netze erläutert. Ferner wird der *P*-CLASSIC-Inferenzalgorithmus vorgestellt und an den Beispielen aus der TV-Domäne eingehend diskutiert.

Inferenzverfahren bilden den Kern eines auf Beschreibungslogiken basierenden Wissenspräsentationssystems. Sie ermöglichen es, anhand des modellierten expliziten Wissens Schlußfolgerungen zu ziehen und neues, implizites Wissen zu gewinnen. Einer der Inferenzdienste sucht z.B. zu dem gegebenen Konzept diejenigen Individuen, die Instanz des Konzepts sind. Es kann auch gefolgert werden, daß ein Konzept von einem anderen Konzept subsumiert wird. In *P*-CLASSIC spricht man von probabilistischer Subsumption, d.h. der Wahrscheinlichkeit, daß eine zufällig gewählte Instanz des Konzepts A auch die des Konzepts B ist. Eine andere Art der Folgerung ist, unter Anwendung eines Verallgemeinerungsoperators (LCS-Operator) aus mehreren Konzepten ein neues Konzept zu generieren. Die Idee ist, das speziellste Konzept zu konstruieren, das die Eingabekonzepte subsumiert. Gerade bei einer beispielorientierten Informationsrecherche kann dieser Operator, genannt „Least Common Subsumer", zum Einsatz kommen.

Betrachten wir hierzu ein Beispiel. Die einzelnen TV-Sendungen wurden in der TV-Wissensbasis als Instanzen von Unterkonzepten eines allgemeinen Konzepts „Sendung" modelliert. Fügt der Benutzer seine Sendungsfavoriten zu einer Sammlung hinzu, ermittelt das System mit Hilfe des LCS-Operators zuerst das speziellste Konzept, welches alle Objekte in der Sammlung beschreibt, und findet anschließend andere Objekte in seiner Wissensbasis, die mit dem berechneten LCS-Konzept auch beschrieben werden können. Diese Objekte werden dem Benutzer als „interessante Sendungen" angeboten. In [18] wurde die LCS-Methode in Verbindung mit CLASSIC eingehend untersucht und zum Teil erweitert. Es wurden auch Grenzen dieser Methode angedeutet. Da bei Bestimmung von LCS-Konzeptkandidaten von der vollständigen Subsumption ausgegangen wird, können die Ergebniskonzepte entweder zu speziell oder aber zu abstrakt sein. Wenn man das o.g. Beispiel betrachtet, stellt man fest, daß die folgenden zwei Grenzfälle möglich sind. Im ersten Grenzfall ist das speziellste Konzept gerade das Oberkonzept „Sendung" und die Menge der „interessanten Sendungen" umfaßt alle Objekte der Wissensbasis. Der zweite Grenzfall liegt vor, wenn das LCS-Konzept so speziell ist, daß keine weiteren Objekte außer den Objekten aus der Sammlung gefunden werden können. Obwohl in beiden Fällen die LCS-Methode theoretisch korrekte Ergebnisse liefert, sind diese Ergebnisse in einer praktischen Anwendung wenig akzeptabel.

Deshalb wurde im Rahmen dieser Arbeit ein neuer Abstraktionsoperator *P*LCS zur Bestimmung von Konzeptgemeinsamkeiten eingeführt, von dem man sich erhofft, die in der Praxis nicht immer wünschenswerten Ergebnisse der LCS-Methode unter Berücksichtigung von Konzeptüberlappungen zu verbessern. Im Kapitel 3 wird das Verfahren zur Berechnung des *P*LCS-Operators ausführlich beschrieben und anhand eines Leitbeispiels aus der TV-Wissensbasis veranschaulicht. Das gesonderte Augenmerk wird auf die Qualitätsmessung der ermittelten *P*LCS-Konzeptkandidaten gerichtet und hierfür im Forschungsfeld „Information Retrieval" verbreitete Bewertungsfaktoren „Recall" und „Precision" herangezogen.

2 Die probabilistische Beschreibungslogik P-Classic

2.1 Vorwort zu Beschreibungslogiken

Beschreibungslogiken, auch als terminologische Logiken bezeichnet, basieren auf der Prädikatenlogik 1. Stufe.

Als ein Basisformalismus für die Wissenspräsentation und Beweisführung liegen sie zahlreichen anwendungsbezogenen KI-Wissenspräsentationssystemen zugrunde. Eins von ihnen ist Classic [2]. In diesem Kapitel wird eine probabilistische Variante – *P*-Classic – vorgestellt.

In der Prädikatenlogik existieren zwei Typen von Ausdrücken: Terme, die Objekte der Welt beschreiben, und Terme, die Beziehungen zwischen Objekten beschreiben. Erstere werden als Konzepte bezeichnet, letztere als Rollen. Aus der Erfahrung mit Datenmodellierungen für praktische Computeranwendungen entstand die Überlegung, daß es sinnvoll ist, gleichartige Objekte zu Gruppen zusammenzufassen, sie zu klassifizieren. Zum Beispiel kann man ein TV-Programm als eine Reihe Sendungen diverser Gattungen betrachten: Filme, Serien, Sportbeiträge, Reportage usw. Konkrete Sendungen wie Tennis-Davis-Cup sind Individuen eines bestimmten Konzepts, beispielsweise des Konzepts Tennissportsendung. Zwischen Individuen existieren zweistellige Relationen, z.B. eine Sportsendung „berichtet über“ eine Sportart.

Angefangen mit dem in der KI mittlerweile zum Quasi-Standard etablierten Wissenspräsentationssystem KL-ONE wurden verschiedene logikorienterte Formalismen ausgearbeitet, die die Idee einer konzeptorienterten Wissensabbildung umsetzen. Dazu gehören auch Beschreibungslogiken, die sich durch die besondere Eigenschaft auszeichnen, eine formale Semantik zu besitzen. Inferenzdienste sind in Bezug auf die formale Semantik definiert.

Zur informellen Einführung einiger im Kontext von Beschreibungslogiken gängiger Begriffe betrachten wir das folgende Beispiel aus der TV-Wissensbasis. Angenommen, man will die Bedingungen formulieren, unter denen ein Objekt als eine „Tennissportsendung“ klassifiziert werden kann. Es gibt nur ein „Sportgerät“ und dieses Gerät ist vom Typ „Tennisschläger“. Das Konzept „Tennissportsendung“ stellt ein Unterkonzept des allgemeinen Konzepts „Sportsendung“ dar. Mit Sprachmitteln von Classic läßt sich diese Beschreibung wie folgt ausdrücken:

```
(define-concept Tennissportsendung
  (and Sportsendung
    (at-least 1 hat-Sportgerät )
    (at-most 1 hat-Sportgerät )
    (all hat-Sportgerät Tennisschläger)))
```

Die Bezeichnungen *Sportsendung* bzw. *Tennisschläger* sind hier Beispiele für die sogenannten primitiven Konzepte. *Primitive Konzepte* benennen Klassen von Individuen und legen notwendige Bedingungen fest, die jede Klasseninstanz erfüllen muß. Konzepte entsprechen den einstelligen Prädikaten der Prädikatenlogik. Die Individuen stehen zu einander in binären Beziehungen, den sogenannten *Rollen*. In dem betrachteten Beispiel verknüpft die Rolle *hat-Sportgerät* das Konzept *Tennissportsendung* mit dem Konzept *Tennisschläger*. Das bedeutet, daß die *Füller* der Rolle *hat-Sportgerät* Individuen des Konzepts *Tennisschläger* sind. Eine Rolle mit nur einem Rollenfüller wird als *Attribut* bezeichnet. Desweiteren kann die Anzahl der Rollenfüller eingeschränkt werden (mit Hilfe der Konstrukte „at-least“ und „at-most“). Das definierte Konzept *Tennissportsendung* bietet ein Beispiel für eine komplexe, aus Basisprimitiven (dazu gehören Konzepte, Rollen, Individuen und termbildende Operatoren) zusammengesetzte Beschreibung. *Definierte Konzepte* postulieren sowohl notwendige als auch hinreichende Bedingungen. Die obige Deklaration wird auch *terminologisches Axiom* genannt. Eine endliche Menge von terminologischen Axiomen wird *Terminologie* bezeichnet.

Die Trennung von terminologischem (Axiome) und assertorischem (Fakten) Wissen ist typisch für Systeme der KL-ONE-Familie. Die terminologische Komponente ist dabei meistens stärker ausgearbeitet.

Es gibt eine Menge von Schlußfolgerungen, die mit Konzepten und Individuen durchgeführt werden können. Die wichtigsten von ihnen sind z.B. Subsumption, Konsistenzüberprüfung, Äquivalenz, Instanzerkennung, um nur einige zu nennen. Insbesondere die Subsumption spielt eine wichtige Rolle. Auf eine formale Spezifikation einiger interessanten Inferenzen in den Beschreibungslogiken wird im Abschnitt 2.3.3 eingegangen.

Konzepte sind aufgrund ihrer Subsumption-Relation zu einander in einer *Taxonomie* (Subsumptionshierarchie) angeordnet. Die speziellsten Konzepte befinden sich unter den allgemeineren Konzepten. Prozesse des Aufsteigens und Absteigens in einer Konzepthierarchie werden als Generalisierung bzw. Spezialisierung bezeichnet. Die auf Beschreibungslogiken basierenden Wissenspräsentationssysteme unterstützen eine automatische Einordnung von Konzepten in eine Subsumptionshierarchie. Daraus ergeben sich viele Vorteile für praktische Anwendungen, die insbesondere bei umfangreichen Informationsbeständen deutlich werden. Beschreibungslogiken finden daher Anwendung in vielen Informationssystemen. Es wurde gezeigt, daß die automatische Klassifikation sehr nützlich sein kann z.B. für die Verwaltung der großen Mengen von Regeln in Expertensystemen [5]. Im Kontext der Informationsrecherche werden gleich mehrere Stärken von Beschreibungslogiken ausgespielt. Beschreibungslogiken eignen sich z.B. zum Formulieren von Anfragen. Eine Anfrage an ein Wissenpräsentationssystem, in dem jedes Objekt eine komplexe Beschreibung hat, ist ebenfalls eine komplexe Beschreibung, die Objekte mit gesuchten Eigenschaften charakterisiert [1]. Beispielsweise kann die Aufforderung „finde alle Sportsendungen, die ausschließlich über Tennis berichten" dann wie folgt ausgedrückt werden:

```
(retrieve
   (and Sportsendung
    (all berichtet-über Tennis)
    (at-least 1 berichtet-über)
    (at-most 1 berichtet-über)))
```

Wird eine Anfrage in der hier beschriebenen deklarativen Weise gestellt, so braucht man nicht auf die speziellen Indexstrukturen zur Verwaltung der gespeicherten Daten Bezug zu nehmen. Die Details der Datenspeicherung bleiben für den Benutzer verborgen.

2.2 Motivation für eine probabilistische Erweiterung

Leider weisen Beschreibungslogiken in folgender Hinsicht Defizite auf. Sie besagen wenig über den Grad der Überlappung von Konzepten. Die angesprochenen Inferenzmechanismen ermöglichen zwar Schlußfolgerungen darüber, ob ein Konzept von anderen subsumiert wird (prinzipielle Überlappung), ob zwei Konzepte disjunkt (keine Überlappung) oder ob sie äquivalent sind (vollständige Überlappung). Sie geben aber nur qualitative Beschreibungen der Konzeptrelationen wieder. Eine quantitative Aussage über das tatsächliche Ausmaß der Überlappung ist mit Beschreibungslogiken nicht möglich.

Das hat zur Folge, daß bei Verwendung von Verallgemeinerungen, die die Ausgangskonzepte vollständig subsumieren, z.B. im Bereich Information Retrieval als Antwort auf Anfragen alle Objekte der Wissensbasis ermittelt werden. Andererseits kann eine zu spezielle Anfrage dazu führen, daß keine neuen Objekte außer der Ausgangsobjekte gefunden werden. Es wird deutlich, daß die Lösung dieses Problems zur Effizienz- und Qualitätsteigerung von Wissenspräsentationssystemen erheblich beitragen kann.

Eine probabilistische Erweiterung der Beschreibungslogik ermöglicht dagegen eine quantitative Aussage über den Grad der Subsumption. In der Wissensbasis für den TV-Assistenten (siehe 2.3.5) wurde z.B. das Wissen postuliert, daß die Konzepte „Individualsport-Sendung" (`individual-sports-broadcast`) und „Mannschaftssport-Sendung" (`team-sports-broadcast`) Unterkonzepte des Konzepts „Sportsendung" (`sports-broadcast`) sind. Es kann aber u.U. sehr wünschenswert sein, die Unter-Ober-Konzeptrelation genauer zu beschreiben, insbesondere wenn diskursspezifische Informationen vorliegen. Beispielsweise kann die Analyse von umfangreichen Datenbeständen ergeben, daß in den meisten Sportsendungen über Mannschaftssportarten berichtet wird. Diese Tatsache würde man als eine Zahl, einen Wahrscheinlichkeitswert festhalten wollen. Zum Beispiel, würde man zum Ausdruck bringen, daß die bedingte Wahrscheinlichkeit, mit der es um eine Mannschaftspielsendung handelt, gleich 0.8 ist, wenn man weiß, daß es eine Sportsendung ist:

$P\,(\texttt{team-sports-broadcast} \mid \texttt{sports-broadcast}) = 0.8$

$P\,(\texttt{individual-sports-broadcast} \mid \texttt{sports-broadcast}) = 0.3$

Andererseits gibt es Sportarten, in denen sowohl Mannschafts- als auch Individualwettkämpfe stattfinden. Diese Information würde man auch als eine statistisch ermittelte Zahl in der Wissensbasis repräsentieren wollen:

$P\,(\texttt{team-sports-broadcast} \mid \texttt{individual-sports-broadcast}) = 0.4$

Einen Lösungsansatz zur Repräsentation von Wissen dieser Art bietet *P*-Classic an, eine probabilistische Erweiterung von Classic. *P*-Classic ermöglicht die Beantwortung folgender Fragen: Mit welcher Wahrscheinlichkeit ist ein Individuum die Instanz eines gegebenen Konzepts? Wie groß ist die Wahrscheinlichkeit, daß eine Instanz eines Konzepts auch die Bedingungen eines anderen Konzepts erfüllt, d.h. mit welcher Wahrscheinlichkeit wird ein Konzept von einem anderen Konzept subsumiert? Oder, um das Beispiel von oben aufzugreifen, was ist die Wahrscheinlichkeit des Konzepts `team-sports-broadcast` unter der Bedingung, daß `individual-sports-broadcast` gilt?

Um die besagte Wahrscheinlichkeit und damit den Grad der Überlappung von Konzepten zu bestimmen, müssen in einem Wissenspräsentationssystem zusätzlich Informationen über die Wahrscheinlichkeitswerte, daß ein Individuum die Instanz eines Konzepts ist, vorhanden sein. Hierfür wird in *P*-Classic eine Reihe sogenannter P-Klassen spezifiziert. Eine P-Klasse beschreibt die Wahrscheinlichkeitsverteilung der Eigenschaften der Individuen in der Anwendungsdomäne.

Die interne Repräsentation einer P-Klasse basiert auf dem Bayes-Netz-Formalismus. Die Grundlagen und die Schlußfolgerungsketten in Bayes-Netzen werden ebenfalls in diesem Teil der Arbeit vorgestellt. Der Inferenzalgorithmus von *P*-Classic setzt auf das Inferenzverfahren von Bayes-Netzen auf. In weiteren Abschnitten dieses Kapitels wird die Kernidee des *P*-Classic-Inferenzverfahrens skizziert und der implementierte Algorithmus anhand der Beispiele detaillierter erläutert.

Die P-Klassen-Abstraktion und das wahrscheinlichkeitstheoretische Inferenzverfahren sind Bestandteile der probabilistischen Komponente von *P*-Classic. Der terminologische Teil von

P-CLASSIC kann als eine eingeschränkte Variante der Beschreibungslogik CLASSIC betrachtet werden. Im folgenden wird zuerst die nicht-probabilistische Komponente von P-CLASSIC formal eingeführt.

2.3 Die terminologische Komponente

2.3.1 Syntax

Eine Sprache zur Beschreibung einer Terminologie besteht aus symbolischen Namen für Konzepte, Rollen und Individuen. Konzeptnamen werden auch als sogenannte atomare Konzepte bezeichnet. Außerdem beinhaltet sie termbildende Konstrukte zum Aufbau von komplexen Beschreibungen. In Abhängigkeit davon, welche Konstrukte unterstützt werden, unterscheiden sich Beschreibungslogiken in ihrer Ausdrucksmächtigkeit. In der Basissprache $\mathcal{AL}$ ist z.B. die Konjunktion ($C \sqcap D$), allgemeingültige Wertbeschränkung für Rollenfüller ($\forall R.C$), unqualifizierte Existenzrestriktion ($\exists R$) sowie $\top$ („Top“, eine Bezeichnung für alle Objekte der Wissensbasis), $\bot$ („Bottom“, ein unerfüllbares Konzept) und Negation von atomaren Konzepten erlaubt. In CLASSIC ist die Sprache $\mathcal{ALN}$ realisiert, in der zusätzlich die Zahlenrestriktionen bzgl. der Anzahl der Rollenfüller zur Verfügung stehen. Eine vollständige Syntax- und Semantik-Spezifikation für Beschreibungslogiken ist z.B. in [15] nachzulesen. Die Tabellen 1 bzw. 2 geben einen Überblick über die Syntax von Konzepten und Ausdrücken in P-CLASSIC. C steht für ein Konzept. A bezeichnet ein atomares Konzept und R ist entweder eine Rolle oder ein Attribut (Q). Ein Individuum ist mit I gekennzeichnet. CN, RN und QN sind Konzept-, Rollen- bzw. Attributnamen. In der ersten Spalte ist die sog. abstrakte Syntax angegeben. In der zweiten Tabellenspalte ist die jeweils äquivalente Darstellung in einer Lisp-Notation verzeichnet, die in folgenden Teilen dieser Arbeit gelegentlich verwendet wird.

$\top$	THING	Top Konzept, das alle Domänenobjekte umfaßt
$\bot$	NOTHING	Unerfüllbares Konzept mit der leeren Extension
$C_1 \sqcap ... \sqcap C_n$	(and $C_1 ... C_n$)	Konjuktion
$\neg A$	(not A)	Negation von atomaren Konzepten
$\forall R.C$	(all R C)	Allgemeingültige Wertbeschränkung
$(\geq\ n\, R)$	(at-least n R)	Minimale Anzahl der Rollenfüller
$(\leq\ n\, R)$	(at-most n R)	Maximale Anzahl der Rollenfüller
$Q : I$	(fills Q I)	Wertzuweisung für Attribute

Tabelle 1: Syntax von Konzepttermen in P-CLASSIC.

$CN \doteq C$	(define-concept CN C)	Definiertes Konzept
$CN \sqsubseteq C$	(define-primitive-concept CN C)	Primitives Konzept
$RN \doteq R$	(define-role RN R)	Definierte Rolle
$RN \sqsubseteq R$	(define-primitive-role RN R)	Primitive Rolle
$AN \doteq R$	(define-attribute AN R)	Definiertes Attribut
$AN \sqsubseteq R$	(define-primitive-attribute AN R)	Primitives Attribut

Tabelle 2: Syntax von Ausdrücken in P-CLASSIC.

$\top$	$\Delta^{\mathcal{I}}$
$\bot$	$\emptyset$
(and $C_1...C_n$)	$C_1^{\mathcal{I}} \sqcap \ldots \sqcap C_n^{\mathcal{I}}$
(not A)	$\Delta^{\mathcal{I}} \backslash A^{\mathcal{I}}$
(all R C)	$\{x \in \Delta^{\mathcal{I}} \vert \forall y \in \Delta^{\mathcal{I}} : (x,y) \in R^{\mathcal{I}} \Rightarrow y \in C^{\mathcal{I}}\}$
(at-least n R)	$\{x \in \Delta^{\mathcal{I}} \vert \sharp\{y \in \Delta^{\mathcal{I}} \vert (x,y) \in R^{\mathcal{I}}\} \geq n\}$
(at-most n R)	$\{x \in \Delta^{\mathcal{I}} \vert \sharp\{y \in \Delta^{\mathcal{I}} \vert (x,y) \in R^{\mathcal{I}}\} \leq n\}$
(fills Q I)	$\{x \in \Delta^{\mathcal{I}} \vert \exists y \in \Delta^{\mathcal{I}} : (x,y) \in R^{\mathcal{I}} \wedge y = I^{\mathcal{I}}\}$

Tabelle 3: Semantik von Konzepttermen in *P*-Classic.

2.3.2 Semantik

Die Semantik des Beschreibungslogik-basierten Teils von *P*-Classic kann mengentheoretisch mit Hilfe einer Interpretation definiert werden. Die Interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ besteht aus einer nichtleeren Menge $\Delta^{\mathcal{I}}$, der sogenannten Domäne von $\mathcal{I}$, und einer Interpretationsfunktion $\cdot^{\mathcal{I}}$, die Konzepte, Rollen, Attribute und Individuen auf ihre Extension in der Interpretation $\mathcal{I}$ abbildet. So wird ein atomares Konzept A auf die Untermenge $A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$, eine atomare Rolle R auf die Untermenge $R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$ (kartesisches Produkt über der Menge $\Delta^{\mathcal{I}}$) und ein Attribut auf eine partielle Funktion von $\Delta^{\mathcal{I}}$ nach $\Delta^{\mathcal{I}}$ abgebildet. Die Erweiterung der Interpretationsfunktion für die Abbildung von komplexeren Konzepten ist in der Tabelle 3 aufgeführt.

Sei A ein atomares Konzept und C ein Konzept in der Beschreibungslogik $\mathcal{ALN}$. Behauptungen der Form $A \sqsubseteq C$ bzw. $A \doteq C$ (als Abkürzung für „$A \sqsubseteq C$ und $C \sqsubseteq A$“) stellen terminologische Axiome dar. Eine endliche Menge terminologischer Axiome wird als eine Terminologie oder T-Box bezeichnet. In der Beschreibungslogik Classic gelten bestimmte Einschränkungen für Axiome aus der T-Box: Kein Konzeptname darf mehr als einmal auf der linken Seite eines terminologischen Axioms vorkommen und selbstbezügliche, d.h. zyklische Axiome sind nicht zulässig. In Classic wird ein atomares Konzept A, für das ein terminologisches Axiom $A \sqsubseteq C$ gilt, als primitives Konzept bezeichnet. Ein Konzept A heißt definiert, wenn ein terminologisches Axiom $A \doteq C$ existiert.

Eine Interpretation $\mathcal{I}$ erfüllt ein terminologisches Axiom $A \sqsubseteq C$ (bzw. $A \doteq C$) genau dann, wenn $A^{\mathcal{I}} \subseteq C^{\mathcal{I}}$ (bzw. $A^{\mathcal{I}} = C^{\mathcal{I}}$) gilt. Eine Interpretation $\mathcal{I}$ ist ein Modell der Terminologie, wenn alle terminologischen Axiome erfüllt sind.

Ein Objekt y ist ein Rollenfüller der Rolle R in Bezug auf ein Objekt x im Modell $\mathcal{I}$, wenn $(x^{\mathcal{I}}, y^{\mathcal{I}}) \in R^{\mathcal{I}}$ gilt.

Die A-Box einer mit einer Beschreibungslogik modellierten Wissensbasis enthält das assertorische Wissen über eine Anwendungsdomäne in Form von assertorischen Axiomen, welche die Behauptungen über die Individuen der Domäne postulieren. Es gibt Konzeptaxiome, die Konzepten Individuen zuweisen (C ist ein Konzept, a ist ein Individuum):

$$C(a)$$

und Rollenaxiome, die Individuen in eine Rolle-Beziehung setzen (R ist eine Rolle, a und b sind Individuen):

$$R(a, b)$$

Eine Interpretation $\mathcal{I}$ erfüllt ein assertorisches Konzeptaxiom genau dann, wenn $a^{\mathcal{I}} \in C^{\mathcal{I}}$

gilt. Ein assertorisches Rollenaxiom wird genau dann von einer Interpretation $\mathcal{I}$ erfüllt, wenn $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in R^{\mathcal{I}}$ gilt.

Eine Interpretation $\mathcal{I}$ ist ein Modell für eine A-Box in Bezug auf eine T-Box, wenn $\mathcal{I}$ ein Modell der T-Box ist und alle assertorischen Axiome in der A-Box erfüllt sind. Auf weitere detaillierte Einführung der A-Box wird an dieser Stelle verzichtet und zum Nachschlagen z.B. auf [20] verwiesen.

Die Wissensbasis in der Beschreibungslogik ist ein Paar $\sum = (T, A)$, wobei T eine T-Box und A eine A-Box kennzeichnen. Eine Interpretation $\mathcal{I}$ ist ein Modell der Wissensbasis $\sum$, wenn sie ein Modell sowohl für T als auch für A ist.

2.3.3 Schlußfolgerungen

Folgend wird ein kurzer Überblick über die wichtigsten Schlußfolgerungen in einer Beschreibungslogik gegeben.

- Konsistenz
  Ein Konzept C gilt als konsistent (erfüllbar) bzgl. einer T-Box genau dann, wenn ein Modell $\mathcal{I}$ der T-Box existiert, so daß $C^{\mathcal{I}} \neq \emptyset$ Gültigkeit hat.

- T-Box-Konsistenz
  Eine T-Box ist konsistent genau dann, wenn für sie ein Modell existiert.

- Subsumption
  Ein Konzept C subsumiert ein Konzept D ($D \preceq C$) in Bezug auf eine T-Box genau dann, wenn $D^{\mathcal{I}} \subseteq C^{\mathcal{I}}$ für alle Modelle $\mathcal{I}$ der T-Box gilt.

- Äquivalenz
  Zwei Konzepte C und D sind äquivalent, oder semantisch gleich ($C \equiv D$) in Bezug auf eine T-Box genau dann, wenn $C^{\mathcal{I}} = D^{\mathcal{I}}$ für alle Modelle $\mathcal{I}$ der T-Box gilt.

- A-Box-Konsistenz
  Eine A-Box ist konsistent in Bezug auf eine T-Box genau dann, wenn es ein Modell $\mathcal{I}$ für die A-Box in Bezug auf die T-Box gibt.

- Instanzerkennung
  Ein Individuum i ist eine Instanz eines Konzepts C genau dann, wenn $i^{\mathcal{I}} \in C^{\mathcal{I}}$ für alle Modelle $\mathcal{I}$ von $\sum$ gilt.

- Retrieval
  Es werden alle Individuen ermittelt, die Instanzen eines Konzepts C sind.

Die Subsumption wird als eine der grundliegenden Relationen angesehen. Die Begründung liegt darin, daß viele andere Schlußfolgerungen sich auf die Subsumption reduzieren lassen. Zum Beispiel kann die Inkonsistenz eines Konzepts C dadurch erkannt werden, daß die Aussage $C \preceq \bot$ gilt. Zwei Konzepte C und D sind disjunkt, wenn $(C \sqcap D) \preceq \bot$ gefolgert wird. Äquivalenz ist die wechselseitige Subsumption: $(C \preceq D) \sqcap (D \preceq C)$.

2.3.4 Die kanonische Form

Für weitere Betrachtungen wird der Begriff „kanonische Form“ benötigt. Ein $\mathcal{ALN}$-Konzept C befindet sich in der kanonischen Form, wenn es wie folgt aussieht:

$$C = \alpha \sqcap \beta_{R_1} \sqcap \ldots \sqcap \beta_{R_k}$$

Das α steht für die Konjunkion von atomaren oder negierten atomaren Konzepten bzw. Wertzuweisungen für Attribute (fills Q a). Es gilt die Bedingung, daß keins der Konzepte oder Attribute in α mehr als einmal vorkommt. Die β_{R_i}-Teile haben die folgende Form: $(\geq\ l_{R_i}\ R_i) \sqcap (\leq\ m_{R_i}\ R_i) \sqcap \forall\, R_i.C_{R_i}$, wobei $R_1, \ldots, R_k$ die Rollen bezeichnen und C_{R_i} wieder Konzepte in der kanonischen Form sind.

Zum Beispiel befindet sich der folgende Konzeptterm in der kanonischen Form:

`Sportsendung` $\sqcap\ \neg$ `Sportgerät` $\sqcap$
$(\leq 1$ `hat-Sportgerät`$)\ \sqcap$
$(\geq 0$ `hat-Sportgerät`$)\ \sqcap$
$\forall$ `hat-Sportgerät.Tennisschläger`

Der Teilausdruck

$(\leq 1$ `hat-Sportgerät`$)\ \sqcap$
$(\geq 0$ `hat-Sportgerät`$)\ \sqcap$
$\forall$ `hat-Sportgerät.Tennisschläger`

stellt einen β-Teil für die Rolle `hat-Sportgerät` dar.

Die *Tiefe* einer Beschreibung in der kanonischen Form wird rekursiv wie folgt bestimmt. Der Teilausdruck α hat die Tiefe 0. Die Tiefe des Konzepts $\alpha \sqcap \beta_{R_1} \sqcap \ldots \sqcap \beta_{R_k}$ berechnet sich als maximaler Wert der Tiefe von Teilausdrücken $\alpha, \beta_{R_1}, \ldots, \beta_{R_k} + 1$.

In dem obigen Beispiel ist die Tiefe des Teilausdrucks `Tennisschläger` gleich 0. Das Konzept C insgesamt hat die Tiefe 1.

2.3.5 Fragment einer Wissensbasis für einen Web-TV-Assistenten

Die Daten einer TV-Domäne wurden mit Mitteln des Wissenspräsentationssystems CLASSIC modelliert. Als Leitbeispiel zur Illustration von *P*-CLASSIC wurde ein kleines Fragment der umfangreichen TV-Wissensbasis ausgewählt, welches sich mit der Repräsentation von Sportsedungen befaßt. Der Oberbegiff für eine allgemeine Sportsendung ist in Form eines atomaren Konzepts `sport-broadcasts` festgelegt. Desweiteren wird zwischen den Sportsendungen unterschieden, die über die Individual- und Mannschaftssportarten berichten (`individual-sports-broadcasts` und `team-sports-broadcast`). Diverse Sportgerättypen werden benannt, z.B. Tennisschläger (das Konzept `tennis-racket`), Fußball (das Konzept `football`), Basketball (das Konzept `basketball`). Sie stellen die Unterkonzepte eines generellen Konzepts für Sportgeräte (`sports-tool`) dar. Sportsendungen und Sportgeräte sind durch die Relation „hat Sportgerät“ (`has-sports-tool`) miteinander in Verbindung gebracht. Aufbauend auf primitiven Konzepten und Rollen sind die Eigenschaften der speziellen Sportsendungen beschrieben. Beispielsweise deklariert das Konzept `football-broadcast` symbolisch notwendige und hinreichende Bedingungen für eine Fußballübertragung. Eine Fußballsportsendung ist eine Mannschaftssportsendung, in der mit genau einem Sportgerät von Typ Fußball gespielt wird. Abbildung 4 gibt einen Überblick über die entstandene Konzepthierarchie.

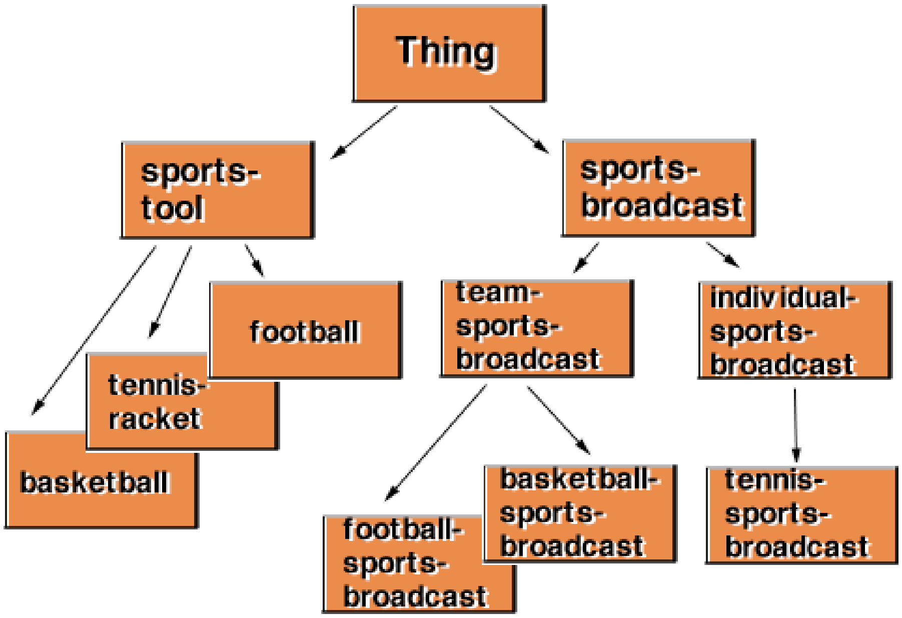

Abbildung 4: Fragment einer TV-Wissensbasis.

Nachfolgend sind die erforderlichen Definitionen in der Classic-Notation aufgeführt (das Konzept `thing` steht für das „top"-Konzept, welches alle Objekte der Anwendungsdomäne beschreibt).

```
(define-primitive-concept sports-broadcast thing)
(define-primitive-concept team-sports-broadcast sports-broadcast)
(define-primitive-concept individual-sports-broadcast sports-broadcast)
(define-primitive-concept sports-tool thing)
(define-primitive-concept basketball sports-tool)
(define-primitive-concept football sports-tool)
(define-primitive-concept tennis-racket sports-tool)
   (define-primitive-role has-sports-tool)
   (define-concept football-sports-broadcast
  (and team-sports-broadcast
    (at-least 1 has-sports-tool)
    (at-most 1 has-sports-tool)
    (all has-sports-tool football)))
   (define-concept basketball-sports-broadcast
  (and team-sports-broadcast
    (at-least 1 has-sports-tool)
    (at-most 1 has-sports-tool)
    (all has-sports-tool basketball)))
   (define-concept tennis-sports-broadcast
  (and individual-sports-broadcast
    (at-least 1 has-sports-tool)
    (at-most 1 has-sports-tool)
```

(`all` *has-sports-tool tennis-racket*)))

2.3.6 Besonderheiten

Auf den ersten Blick scheint die terminologische Komponente von *P*-CLASSIC sich weitgehend an die Konventionen von CLASSIC zu orientieren. Es gibt allerdings eine Reihe Einschränkungen bzw. Erweiterungen in Bezug auf die Beschreibungslogik CLASSIC, die an dieser Stelle erwähnt werden sollen.

Die Anzahl der Rollenfüller ist in *P*-CLASSIC begrenzt. Diese Restriktion ist aus zwei Überlegungen entstanden. Es wird im Abschnitt 2.4.7 gezeigt, daß der Inferenzalgorithmus von *P*-CLASSIC alle möglichen Werte der Anzahl der Rollenfüller berücksichtigt. Ohne die Einschränkung der maximalen Anzahl von Rollenfüllern würde er nie terminieren. Zum anderen muß sichergestellt werden, daß die Wahrscheinlichkeitsverteilung für alle Kombinationen der Anzahl der Rollenfüller vollständig spezifiziert werden kann. In [7] wurden einige Ansatzpunkte zur Aufhebung dieser Restriktion diskutiert (z.B. könnte die Wahrscheinlichkeit, daß es n Rollenfüller gibt, mit Hilfe einer hypergeometrischen Funktion f(n) ausgedrückt werden). Im Rahmen dieser Arbeit wurde dieses Problem nicht näher betrachtet.

Die Sprache von *P*-CLASSIC bietet direkt keine Konstrukte an, um die Disjunktheit bzw. Subsumption von Konzepten zu deklarieren. Aus der nachfolgenden Diskussion wird dennoch ersichtlich, daß die probabilistische Komponente den Ausdruck von Beziehungen dieser Art zwischen den atomaren Konzepten doch erlaubt. Über die Subsumptionsrelation von atomaren Konzepten kann mit Hilfe der Wahrscheinlichkeitsberechnungen geschlossen werden. Beispielsweise läßt sich die Aussage treffen, ob das Konzept „Mannschaftssportsendung" vom Konzept „Sportsendung" subsumiert wird (`team-sports-broadcast` $\preceq$ `sports-broadcast`). In der Tat, wenn man berücksichtigt, daß die Subsumptionsrelation zwischen zwei Konzepten $C \preceq D$ auch wie folgt umgeschrieben werden kann: $C \sqcap \neg D$, dann genügt es zu zeigen, daß die Wahrscheinlichkeit beider Ereignisse „C gilt" und „$\neg D$ gilt" gleich 0 ist, um zu schließen, daß D das Konzept C subsumiert. Eine wichtige Voraussetzung dafür ist allerdings, daß *P*-CLASSIC im Gegensatz zu CLASSIC die Negation von atomaren Konzepten unterstützt. Durch die Einführung in die P-Klassen zusätzlicher Knoten für negierte atomare Konzepte sind Schlußfolgerungen mit diesen Konzepte möglich. Details werden in den folgenden Kapiteln erläutert. Unter bestimmten Voraussetzungen können sogar Schlüsse über die Subsumption von komplexen Konzepttermen gezogen werden. Zum Beispiel kann ermittelt werden, ob die Aussage

`tennis-sports-broadcast` $\preceq$
`sports-broadcast` $\sqcap \forall$ `has-sports-tool` . `sports-tool`

gilt. Dafür müssen die konketen Wahrscheinlichkeitsverteilungen für Anwendungsobjekte analysiert werden. So genügt es zu zeigen, daß die Wahrscheinlichkeit von `tennis-sports-broadcast` nur dann größer als 0 sein kann, wenn das Konzept `sports-broadcast` gilt. Sollte das Konzept `sports-broadcast` gelten, so genügt es zu zeigen, daß die Füller der Rolle `has-sports-tool` nur vom Typ `sports-tool` sein können. Für die ricthige Auswahl der Wahrscheinlichkeitsverteilung (P-Klasse) der Rollenfüller einer Rolle sorgt ein spezieller Knoten. Und wenn in der P-Klasse für Rollenfüller vom Typ `sports-tool` das Konzept `sports-tool` die A-priori-Wahrschein-lichkeit 1 hat, dann ist die Gültigkeit der ganzen Aussage bewiesen.

Nachdem der terminologische Teil von *P*-CLASSIC diskutiert wurde, geht der folgende Abschnitt zur Beschreibung des zweiten *P*-CLASSIC-Bestandteils über. Die Semantik, die P-

Klassen und das Inferenzverfahren der probabilistischen Komponente werden im folgenden thematisiert und anhand von Beispielen veranschaulicht.

2.4 Die probabilistische Komponente

Die probabilistische *P*-Classic-Komponente wurde mit dem Ziel entwickelt, eine Wissensbasis um die probabilistischen Axiome und Behauptungen zu erweitern. Wäh-rend in der T-Box der terminologischen Komponente von *P*-Classic das allgemeine Wissen über die Beziehungen von Konzepten modelliert werden kann, sind mit Hilfe einer probabilistischen Erweiterung zusätzlich quantitative Aussagen über die Subsumptionsrelation von Konzepten möglich. Die wichtigste Schlußfolgerung ist dabei die probabilistische Subsumption: Gegeben sind zwei Konzepte C und D. Wie groß ist die Wahrscheinlichkeit $P(C|D)$, daß für ein Individuum des Konzepts D auch die Bedingungen des Konzepts C erfüllt sind? Die Größe der Wahrscheinlichkeit kann selbstverständlich 0 (keine Subsumption) oder 1 (vollständige Subsumption) sein. Damit sind die Schlußfolgerungen über die Konzeptsubsumption bzw. Disjunktheit, wie sie in der Inferenzkomponente des terminologischen Teils von *P*-Classic realisiert sind, weiterhin möglich. Das Interenzverfahren von *P*-Classic ermöglicht allerdings eine weitere Verfeinerung des Grades der Subsumption, so daß die Größe der Konzeptüberlappung als eine reelle Zahl aus dem Wertebereich 0 bis 1 dargestellt werden kann.

Damit die Beantwortung von probabilistischen Anfragen realisiert werden kann, muß die Wissensbasis die Wahrscheinlichkeitswerte der logischen Aussagen enthalten. Dabei ist folgendes zu beachten. Wenn ein Sachverhalt durch n Aussagen beschrieben ist, müssen die Wahrscheinlichkeiten von 2^n Kombinationen der Aussagen bekannt sein, um die gemeinsame Wahrscheinlichkeitsverteilung zu berechnen. Für n größer 10 ist die Komplexität der Speicherung und der Berechnung von Rand- und bedingten Wahrscheinlichkeiten nicht beherrschbar. Abgesehen davon sind die gefordeten Informationen auch nicht verfügbar, da es in praktischen Anwendungen beinah unmöglich ist, über die Zusammenhänge von allen Ereignissen untereinander eine Aussage zu machen [11].

Dennoch reicht unter der Voraussetzung der bedingten Unabhängigkeiten der logischen Aussagen, wie im folgenden erläutert wird, eine wesentlich kleinere Datenmenge aus, um die gegenseitigen Abhängigkeiten von n Aussagen zu beschreiben. Unter diesen Bedingungen können die Wahrscheinlichkeitsaussagen in Form von Bayes-Netzen dargestellt und verarbeitet werden. Die probabilistische Komponente von *P*-Classic besteht aus gleich mehreren solchen Bayes-Netzen (P-Klassen), welche die Wahrscheinlichkeitsverteilung der Eigenschaften der Anwendungsobjekte abbilden. Es mag auf den ersten Blick merkwürdig erscheinen, daß mehrere Wahrscheinlichkeitsverteilungen notwendig sind. Es darf dabei nur nicht der folgende Umstand übersehen werden. Wenn bekannt ist, daß ein Anwendungsobjekt auch als Rollenfüller bzgl. eines anderen Objektes auftritt, dann muß man tatsächlich eine andere Wahrscheinlichkeitsverteilung für die Konzeptzugehörigkeit dieses Objektes zugrundelegen. Zum Beispiel haben Raubtiere, die jagen, andere Wahrscheinlichkeitsverteilung, als Raubtiere, die selbst gejagt werden.

Um die Bedeutung von P-Klassen zu verdeutlichen, muß man auf die Grundbegriffe der Wahrscheinlichkeitstheorie zurückgreifen. Im folgenden werden deshalb als erstes einige für das Verständnis relevante Begriffe und Formeln der Wahrscheinlichkeitsrechnung aufgeführt und der Bayes-Netz-Formalismus vorgestellt. Die Intention dieses Abschnitts besteht darin, die probabilistische Komponente in ihrer Semantik zu beschreiben und einen Inferenzalgorithmus zu präsentieren.

2.4.1 Einige Grundbegriffe der Wahrscheinlichkeitstheorie

In der Wahrscheinlichkeitsrechnung werden primär die Ereignisse betrachtet, deren Eintreten unvorhersehbar ist [1]. In einem stochastischen Experiment wird ein Stichprobenraum $\Omega = \{e_1, e_2, \ldots, e_m\}$ aufgespannt, der die Menge aller möglichen Ergebnisse eines Experiments, genannt *Elementarereignisse*, umschließt. Jede Teilmenge der Menge von Elementarereignissen $A \subseteq \Omega$ stellt ein *zufälliges Ereignis* dar. Die Potenzmenge 2^Ω ist die Menge aller möglichen Ereignisse und wird als *Ereignisraum* bezeichnet. Ereignisverknüpfungen können mit Hilfe von Mengenoperationen definiert werden. Die Summe von Ereignissen wird als $A_1 \cup A_2$ und das Produkt als $A_1 \cap A_2$ bezeichnet. Zwei Ereignisse A_1 und A_2 heißen *unvereinbar*, wenn sie nicht gemeinsam auftreten können, d.h. $A_1 \cap A_2 = \emptyset$ gilt. Für eine beliebige Menge von Ereignissen $\{A_1, \ldots, A_m\}$ besteht ein vollständiges System unvereinbarer Ereignisse aus 2^m Produkten, in denen jedes Ereignis oder dessen Komplement ($\neg A = \Omega \setminus A$) vorkommt.

Die Auswertung der Ergebnisse eines Experimentes führt zur Bestimmung der relativen Häufigkeit, mit der ein Ereignis auftritt. Der Grenzwert der relativen Häufigkeit des Auftretens eines Ereignisses bei unendlich vielen Stichproben bezeichnet man als *Wahrscheinlichkeit* eines Ereignisses. In der Wissensverarbeitung werden die Ergebnisse eines Experiments $e_1, e_2, \ldots, e_m$ durch logische Aussagen $p_1, p_2, \ldots, p_m$ beschrieben. Ein Ereignis e_i ist genau dann eingetreten, wenn die entsprechende logische Aussage p_i den Wert „wahr“ hat, während alle anderen Aussagen falsch sind.

Die *bedingte Wahrscheinlichkeit* $P(A_1|A_2)$ gibt an, wie groß die Wahrscheinlichkeit des Ereignisses A_1 ist, wenn bekannt ist, daß das Ereignis A_2 aufgetreten ist.

Es gilt:

$$P(A_1|A_2) = \frac{P(A_1 \cap A_2)}{P(A_2)}$$

Anhand dieser Formel ist die Wahrscheinlichkeit des gemeinsamen Auftretens von A_1 und A_2 folgendermaßen zu berechnen:

$$P(A_1 \cap A_2) = P(A_1|A_2) \cdot P(A_2)$$

Erweitert auf n Ereignisse folgt daraus die sogenannte *Kettenregel* der bedingten Wahrscheinlichkeiten, die die Wahrscheinlichkeit des gemeinsamen Auftretens von n Ereignissen als Produkt von einer „Kette“ der bedingten Wahrscheinlichkeiten darstellt:

$$\begin{aligned} &P(A_1 \cap A_2 \cap \ldots \cap A_n) = \\ &\quad P(A_1|A_2 \cap \ldots \cap A_n) \cdot P(A_2|A_3 \cap \ldots \cap A_n) \ldots P(A_{n-1}|A_n) \cdot P(A_n) \end{aligned}$$

Wenn alle Ereignisse voneinander unabhängig sind, vereinfacht sich die Kettenregel zu der folgenden Formel:

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1) \cdot P(A_2) \ldots P(A_n)$$

Die Wahrscheinlichkeit der Summe von Ereignissen läßt sich wie folgt berechnen:

$$\begin{aligned} P(A_1 \cup A_2 \cup \ldots \cup A_n) &= \sum_{k=1,\ldots,n} P(A_k) - \sum_{k_1<k_2} P(A_{k_1} \cap A_{k_2}) + \\ &\quad \sum_{k_1<k_2<k_3} P(A_{k_1} \cap A_{k_2} \cap A_{k_3}) + \ldots \\ &\quad \ldots + (-1)^{n+1} P(A_1 \cap A_2 \cap A_n) \end{aligned}$$

[1] Die in diesem Abschnitt verwendeten Definitionen stammen weitgehend aus [11].

2.4.2 Noch mehr Formeln aus der Wahrscheinlichkeitstheorie

Die Ergebnisse eines zufälligen Ereignisses werden oft einer Variable X zugewiesen. Da es im allgemeinen um mehrere Werte handelt, stellt X einen Vektor dar. Die *Zufallsvariable* X ist die Abbildung der Elementarereignisse Ω auf den Wertebereich W_X:

$$X : \Omega \to W_x$$

Die Wahrscheinlichkeit mehrerer Ereignisse A_1, A_2, …, A_m läßt sich als Funktion

$$P(X_1, X_2, \ldots, X_m) : W_{X_1} \times W_{X_2} \times \ldots \times W_{X_m} \to [0, 1]$$

auffassen, wobei $X_1, X_2, \ldots, X_m$ Zufallsvariablen sind. Diese Funktion wird als *Wahrscheinlichkeitsverteilung* bezeichnet (m-dimensional).

Die Summe der Wahrscheinlichkeiten, mit der die erste Zufallsvariable den Wert X_1, die zweite Variable den Wert X_2 usw. annimmt, ist gleich 1:

$$\left(\sum_{X_1 \in W_{X_1}, \ldots, X_m \in W_{X_m}} P(X_1, X_2, \ldots, X_m) \right) = 1$$

Die *Randwahrscheinlichkeitsverteilung* beschriebt die Wahrscheinlichkeitsverteilung einer bzw. mehrerer Variablen X_i unabhängig davon, welchen Wert die anderen Variablen haben.

Zum Beispiel läßt sich die Randwahrscheinlichkeitsverteilung der zufälligen Variable X_1 als Summe über alle Wertkombinationen der nicht auf der linken Seite der Gleichung stehenden Variablen $X_2, \ldots, X_m$ ausrechnen:

$$P(X_1) = \sum_{X_2 \in W_{X_2}, \ldots, X_m \in W_{X_m}} P(X_1, X_2, \ldots, X_m)$$

Die Kettenregel führt zu der folgenden Formel zur Berechnung der *gemeinsamen Wahrscheinlichkeitsverteilung* der Zufallsvariablen $X_1, X_2, \ldots, X_m$:

$$\begin{aligned} P(X_1, X_2, \ldots, X_m) &= P_{X_1|X_2,\ldots,X_m}(X_1, X_2, \ldots, X_m) \\ &\quad P_{X_2|X_3,\ldots,X_m}(X_2, X_3, \ldots, X_m) \ldots \\ &\quad P_{X_{m-1}|X_m}(X_{m-1}, X_m) P_{X_m}(X_m) \end{aligned}$$

Bei der aussagenlogischen Darstellung entspricht der Wert der Zufallsvariable X_i dem Wahrheitswert der Aussage p_i für das Ergebnis des betrachteten Experiments. Wenn ein Problem durch n Aussagen beschrieben wird, ergibt sich die gemeinsame Wahrscheinlichkeitsverteilung durch die Angabe von 2^n Werten. Bei 2 Aussagen erhält man z.B. das folgende wahrscheinlichkeitstheoretische Modell:

$$\begin{aligned} P_{X_1,X_2}(T, T) &= P(p_1 \wedge p_2) \\ P_{X_1,X_2}(T, F) &= P(p_1 \wedge \neg p_2) \\ P_{X_1,X_2}(F, T) &= P(\neg p_1 \wedge p_2) \\ P_{X_1,X_2}(F, F) &= P(\neg p_1 \wedge \neg p_2) \end{aligned}$$

Anhand der gemeinsamen Wahrscheinlichkeitsverteilung können beliebige bedingte und Randwarscheinlichkeitsverteilungen für einzelne oder mehrere Zufallsereignisse berechnet werden. Generell werden solche Schritte als *probabilistische Inferenzen* bezeichnet.

Für große Werte von n stellt die Berechnung und Speicherung der gemeinsamen Wahrscheinlichkeitsverteilung ein Komplexitätsproblem dar. Für praktische Anwendungen ist die Kenntnis aller bedingten Wahrscheinlichkeiten meistens auch nicht erforderlich. Deshalb wurden Bedingungen formuliert, die dazu führen, daß eine wesentlich kleinere Datenmenge ausreicht, um die gegenseitigen Abhängigkeiten der n Aussagen vollständig zu beschreiben.

Betrachten wir ein Problem, welches durch z.B. 3 Aussagen p_1, p_2 und p_3 beschrieben ist. Für ein vollständiges wahrscheinlichkeitstheoretisches Modell müssen 8 Werte der gemeinsamen Wahrscheinlichkeitsverteilung bekannt sein. Man schränkt die gegenseitigen Abhängigkeiten der Aussagen so ein, daß es eine Kausalkette entsteht: $p_1 \to p_2 \to p_3$. Ist die Aussage p_1 wahr, so kann mit bestimmter Wahrscheinlichkeit auf die Gültigkeit von p_2 geschlossen werden. Gilt p_2, dann folgt daraus p_3. Wenn den Aussagen die Zufallsvariablen x_1, x_2 und x_3 zugeordnet sind, läßt sich die Wahrscheinlichkeitsverteilung für p_3 unter Anwendung der Kettenregel ableiten:

$$P(x_1, x_2, x_3) = P_{x_3|x_1,x_2}(x_1, x_2, x_3)P_{x_2|x_1}(x_1, x_2)P_{x_1}(x_1)$$

Da alle drei Aussagen gelten, ist $x_1 = T, x_2 = T$ und $x_3 = T$. Die Wahrscheinlichkeit des gemeinsamen Auftretens von p_1, p_2 und p_3 ist ein Produkt der bedingten Wahrscheinlichkeiten:

$$P(p_1 \wedge p_2 \wedge p_3) = P(p_3|p_1 \wedge p_2)P(p_2|p_1)P(p_1)$$

Es wird deutlich, daß die Wahrscheinlichkeit des gemeinsamen Auftretens von allen 3 Aussagen im allgemeinen nicht allein aus der Kenntnis der bedingten Wahrscheinlichkeiten $P(p_2|p_1)$ und $P(p_3|p_2)$ berechnet werden kann, sondern auch die Kenntnis der Wahrscheinlichkeit des Ereignisses p_3 unter der Bedingung, daß p_1 und p_2 aufgetreten sind ($P(p_3|p_1 \wedge p_2)$), notwendig ist. Soll die Wahrscheinlichkeit von p_3 doch allein anhand der bedingten Wahrscheinlichkeiten $P(p_2|p_1)$ und $P(p_3|p_2)$ berechenbar sein, muß eine zusätzliche Bedingung festgelegt werden – die bedingte Unabhängigkeit. Wenn nämlich bekannt ist, daß die Aussage p_2 gültig ist, dann ist die Wahrscheinlichkeit von p_3 nicht mehr von p_1 abhängig. Man sagt, p_2 „blockiert" die stochastische Abhängigkeit zwischen p_1 und p_3. Anders ausgedrückt, das Ereignis p_2 „separiert" Ereignisse p_1 und p_3.

Unter der Voraussetzung, daß die Aussagen bedingt unabhängig voneinander sind, können Schlußfolgerungsketten aufgebaut werden. Die gemeinsame Wahrscheinlichkeitsverteilung wird schrittweise berechnet. Es wird bei dem ersten Ereignis der Kausalkette angefangen und in Richtung der Kanten, die kausale Abhängigkeiten darstellen, vorangegangen.

$$X_1 \stackrel{P_{X_2|X_1}(X_1,X_2)}{\longrightarrow} X_2 \stackrel{P_{X_3|X_2}(X_2,X_3)}{\longrightarrow} X_3$$

1. P_{X_1}
2. $P_{X_1,X_2} = P_{X_2|X_1}P_{X_1}$
3. $P_{X_1,X_2,X_3} = P_{X_3|X_2}P_{X_2|X_1}P_{X_1} = P_{X_3|X_2}P_{X_1,X_2}$

Im allgemeinen läßt sich die gemeinsame Wahrscheinlichkeitsverteilung dann wie folgt ausrechnen.

$$P_{X_1,X_2,\ldots,X_m} = \prod_{i=1}^{m} P_{X_i|V(X_i)}$$

$V(X_i)$ bezeichnet die Menge der Variablen, von denen X_i abhängt.

Die bedingte Unabhängigkeit der Aussagen und die darauf basierende Inferenzregel sind die Grundlagen für die Bildung von Bayes-Netzen.

2.4.3 Bayes-Netze

Definition Ein Bayes-Netz, auch als „Abhängigkeitsgraph“ oder „belief network“ bezeichnet, ist ein gerichteter azyklischer Graph, dessen Knoten Zufallsvariablen darstellen und dessen Kanten direkte stochastische Abhängigkeiten repräsentieren. Die Gesamtmenge der Zufallsvariablen stellt eine Problemdomäne dar. In Abbildung 5 ist zur Veranschaulichung ein Bayes-Netz mit 4 Knoten dargestellt. Die Knoten sind mit den Zufallsvariablen X_1, X_2, X_3 und X_4 assoziiert. Die Richtung der Pfeile gibt die kausale Abhängigkeit der Knoten wieder.

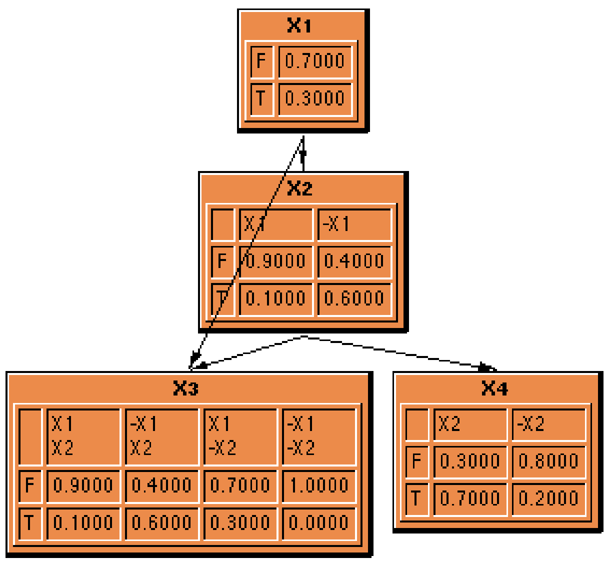

Abbildung 5: Ein einfaches Bayes-Netz.

Der Zweck eines Bayes-Netzes ist, die gemeinsame Wahrscheinlichkeitsverteilung für die Problemdomäne in einer einfachen und übersichtlichen Art und Weise zu modellieren. Das Modell besteht aus einer Menge von „lokalen“ bedingten Wahrscheinlichkeitsverteilungen und basiert auf der Annahme der bedingten Unabhängigkeit der Variablen. Das heißt, daß jede Zufallsvariable den Wert aus einem bestimmten Wertebereich annimmt und genau von denjenigen anderen Variablen bedingt abhängig ist, mit deren Knoten sie durch eine Kante direkt verbunden ist. Die Kanten sind mit den bedingten Wahrscheinlichkeiten der benachbarten Knoten gewichtet.

Conditional Probability Table Die bedingten Wahrscheinlichkeiten der Zufallsvariablen sind in einer sogenannten „Conditional Probability Table“ (CPT) zusammengefaßt.

Eine CPT legt für jeden Knoten des Bayes-Netzes die Wahrscheinlichkeit fest, mit der dieser Knoten abhängig von der Kombination der Werte seiner Vorgängerknoten einen bestimmten Wert annimmt. Der Wurzelknoten ist immer mit einer bestimmten A-priori-Wahrscheinlichkeit gewählt.

In dem obigen Beispiel beinhaltet die CPT des Knotens X_3 8 Werte der bedingten Wahrscheinlichkeiten $P(X_3|X_1)$ und $P(X_3|X_2)$ für alle Wertekombinationen der Knoten X_1, X_2 und X_3:

$$\begin{aligned} P(X_3 = T|X_1 = T, X_2 = T) &= 0.1 \\ P(X_3 = T|X_1 = T, X_2 = F) &= 0.3 \\ &\dots \\ P(X_3 = F|X_1 = F, X_2 = F) &= 0.0 \end{aligned}$$

Unter Berücksichtigung der in einer CPT abgelegten statistischen Informationen können diverse gemeinsame Wahrscheinlichkeiten und Randwahrscheinlichkeiten der Zufallsereignisse berechnet werden.

Probabilistische Inferenzen Die probabilistischen Inferenzen in Bayes-Netzen basieren auf dem zum Schluß des vorangehenden Abschnitts erläuterten Verfahren zum Berechnen der gemeinsamen Wahrscheinlichkeitsverteilung:

$$P_{X_1,X_2,\dots,X_m} = \prod_{i=1}^{m} P_{X_i|V(X_i)}$$

Anhand dieser Formel ist für jede Teilmenge der Zufallsvariablen deren gemeinsame Wahrscheinlichkeitsverteilung in einem Bayes-Netz eindeutig festgelegt und kann direkt aus dem Netz ermittelt werden.

Die gemeinsame Wahrscheinlichkeitsverteilung aller Zufallsvariablen läßt sich in einem Bayes-Netz schrittweise durch das Voranschreiten entlang der Kanten in der mit Pfeilen gezeigten Richtung ableiten. Der Wurzelknoten erhält einen A-priori-Wert. Den „Fluß“ durch das in Abbildung 5 dargestellte Netz veranschaulicht die folgende Schlußfolgerungskette:

- P_{X_1}
- $P_{X_1,X_2} = P(X_2|X_1) \cdot P_{X_1}$
- $P_{X_1,X_2,X_3} = P(X_3|X_1, X_2) \cdot P(X_2|X_1) \cdot P(X_1)$
- $P_{X_1,X_2,X_3,X_4} = P(X_4|P(X_2) \cdot P(X_3|X_1, X_2) \cdot P(X_2|X_1) \cdot P(X_1)$

Anhand der aufgeführten Berechnungsvorschriften können die Wahrscheinlichkeiten logischer Aussagen ermittelt werden. Hier sind nur zwei Beispiele:

$$\begin{aligned} &P(X_1 = T, X_2 = T, X_3 = F) = \\ &\quad P(X_3 = F|X_1 = T, X_2 = T) \cdot P(X_2 = T|X_1 = T) \cdot P(X_1 = T) = \\ &\quad 0.9 \cdot 0.1 \cdot 0.3 = 0.027 \end{aligned}$$

Wenn über die Werte von Vorgängerknoten nichts bekannt ist, greift die Berechungsvorschrift für die Randwahrscheinlichkeiten, wobei alle möglichen Wertkombinationen der Vorgängerknoten berücksichtigt werden müssen:

$$
\begin{aligned}
&P(X_2 = T) = \\
&\quad P(X_1 = T, X_2 = T) + P(X_1 = F, X_2 = T) = \\
&\quad P(X_2 = T | X_1 = T) \cdot P(X_1 = T) + P(X_2 = T | X_1 = F) \cdot P(X_1 = F) = \\
&\quad 0.1 \cdot 0.3 + 0.6 \cdot 0.7 = 0.45
\end{aligned}
$$

Das beschriebene Verfahren stellt natürlich eine „Brute-force"-Methode dar. Um die Kosten einer solchen Prozedur zu reduzieren, wurden diverse optimierende Verfahren ausgearbeitet. Zum Beispiel wurde von Lauritzen und Spiegelhalter [9] ein Algorithmus entwickelt, der das Bayes-Netz zu einem ungerichteten Graph transformiert und die mathematischen Eigenschaften von ungerichteten Graphen ausnutzt. In Arbeiten von B. D'Ambrosio und Z. Li [10] wurde ein Verfahren vorgeschlagen, in dem die Summen und Produkte der bedingten Wahrscheinlichkeiten in der Inferenzkette symbolisch vereinfacht werden können. In Forschungen von Pearl wurde ein Szenario entworfen, das auf der Propagierung in Poly-Bäumen basiert [16]. Eine der naheliegenden Lösungen ist das „Caching" und Wiederverwenden von in den einzelnen Schritten der Schlußfolgerungskette ermittelten Wahrscheinlichkeiten. Die Implementation des *P*-Classic-Inferenzverfahrens greift im ersten Prototyp auf die erwähnte „Brute-force"-Methode zurück und unterstützt nur ansatzweise das Caching. In [7] wurde gezeigt, daß Inferenzen in *P*-Classic in polynomialer Zeit berechnen werden können, sobald die Berechnung der probabilistischen Inferenzen in Bayes-Netzen auch in polynomialer Zeit erfolgt. Das wäre eine Motivation für die künftige Weiterentwicklung des *P*-Classic-Rahmensystems in der Richtung einer Verknüpfung des *P*-Classic-Inferenzalgorithmus mit Optimierungstechniken in Bayes-Netzen.

D-Separation Die Topologie der Bayes-Netze bildet die Abhängigkeitsrelationen zwischen den Zufallsereignissen ab. Nicht nur die Existenz, sondern auch die Richtung der Kanten in einem Bayes-Netz spielt dabei eine wichtige Rolle. Sind die Knoten miteinander durch Kanten verbunden, so heißt es nicht automatisch, daß sie voneinander abhängen. Zur Veranschaulichung wird auf das Beispiel eines einfachen Bayes-Netzes zurückgegriffen (siehe Abbildung 6). Es gibt zwar einen Weg von X_2 über X_4 zu X_3. Die Richtung der Kanten auf diesem Weg ist aber von entscheidender Bedeutung. Die Abhängigkeit zwischen X_2 und X_3 ist solange „blockiert", d.h. X_2 und X_3 sind solange als unabhängige Zufallsereignisse zu betrachten, bis das Ereignis X_4 eingetreten ist und die mit ihm assoziierte Zufallsvariable einen bestimmten Wert angenommen hat. Erst das Initiieren der Zufallsvariable X_4 bewirkt die Abhängigkeitsrelation zwischen X_2 und X_3. Man sagt, daß die Menge $Z_1 = \{X_4\}$ X_2 von X_3 „d-separiert".

Das Kriterium für die Konnektivität der Knoten in einem gerichteten azyklischen Graph wird als „d-Separation" bezeichnet. Dieses Kriterium ermöglicht es, anhand von formalen Testbedingungen, die im folgenden erläutert werden, zu bestimmen, ob zwei Mengen der Zufallsvariablen in Bezug auf eine dritte Menge voneinander unabhängig sind. Die folgende Definition stammt aus [17].

Seien V, W und Z drei disjunkte Teilmengen einer Menge von Knoten in einem DAG. Die Menge Z „d-separiert" V von W, wenn auf jedem Pfad zwischen einem beliebigen Knoten aus

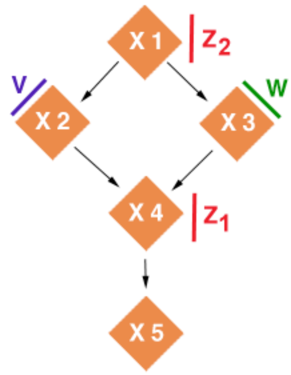

Abbildung 6: D-Separation.

V und einem beliebigen Knoten aus W ein Knoten w liegt, für den eine der beiden folgenden Bedingungen gilt:

1. Der Knoten w hat eingehende Pfeile und weder w noch die Nachfolgerknoten von w sind in Z.

2. Keine Pfeile gehen in den Knoten w ein und w ist in Z.

Soll ein Pfad die angegebenen Voraussetzungen erfüllen, so wird er als „blockiert durch Z“ gekennzeichnet. Andererfalls wird er als „aktiv“ genannt.

Am Beispiel eines DAG in Abbildung 6 kann jetzt gezeigt werden, daß die Menge V=$\{X_2\}$ und W=$\{X_3\}$ auch durch die Knotenmenge $Z_2 = \{X_1\}$ d-separiert sind. In der Tat ist der erste Pfad von X_2 zu X_3 ($X_2 \longleftarrow X_1 \longrightarrow X_3$) durch X_1 blockiert. Der zweite mögliche Pfad $X_2 \longrightarrow X_4 \longleftarrow X_3$ ist ebenfalls blockiert, weil der einzige Knoten mit eingehenden Pfeilen X_4 und sein Nachfolger X_5 nicht in Z_2 enthalten sind.

Das hier eingeführte Kriterium „d-Separation“ wird bei der Definition von P-Klassen verwendet.

2.4.4 P-Klassen

Nachdem die zum Verständnis von P-Klassen notwendigen Grundlagen erläutert wurden, bietet dieser Abschnitt eine eingehende Diskussion über die P-Klassen an.

Eine P-Klasse deklariert die Wahrscheinlichkeitsverteilung der Eigenschaften der Objekte aus einer Anwendungsdomäne $\Delta^{\mathcal{I}}$ und ist als Bayes-Netz modelliert. Aus der wahrscheinlichkeitstheoretischen Perspektive kann eine P-Klasse auch wie folgt aufgefaßt werden. Sie beschreibt ein zufälliges Ereignis, daß bei einem Experiment (im Sinne der Wahrscheinlichkeitsrechnung) ein Individuum die Instanz eines Konzepts ist. Ein solches Ereignis ist genau dann eingetreten, wenn die logische Aussage, welche die Eigenschaften des Individuums beschreibt, gilt. Was sind nun die Eigenschaften eines Individuums? Wie der Diskussion über die T-Box in 2.3.2 entnommen werden kann, sind die Grundeigenschaften eines Individuums in Form von atomaren Konzepten festgelegt. Ist das Ereignis aufgetreten, sind die Konzepte erfüllt. Desweiteren sind den Attributen des Individuums bestimmte Werte zugewiesen und die Anzahl der Rollenfüller ist für jede Rolle eingeschränkt. Die Wahl der Rollenfüller selbst stellt schon ein anderes zufälliges Ereignis dar, welches durch eine eigene Wahrscheinlichkeitsverteilung charakterisiert ist. Auf diesen wichtigen Punkt muß gesondert hingewiesen werden. Die Wahrscheinlichkeitsverteilung der Individuumeigenschaften ist im allgemeinen nicht dieselbe wie die von Rollenfüllern, auch wenn es u.U. um ein und dasselbe Objekt handelt. Wichtig bleibt, ob das Objekt ein Rollenfüller ist oder nicht, d.h. es ist entscheidend, auf welcher Seite der zweistelligen Rollenbeziehung es vorkommt. Dann müssen nämlich je nach Anzahl der Rollen, für die das Objekt ein Rollenfüller ist, mehrere Wahrscheinlichkeitsverteilungen festgelegt werden. Aus diesem Grund wird jeder Rolle eine oder mehrere P-Klassen zur Beschreibung der Wahrscheinlichkeitsverteilung der Rollenfüller zugeordnet. Vorausgesetzt wird, daß die Wahrscheinlichkeitsverteilungen der Rollenfüller untereinander unabhängig sind. Mit anderen Worten, sowohl die Wahl jedes einzelnen Rollenfüllers als auch die Wahl eines Objektes und seiner Rollenfüller müssen unabhängige Ereignisse darstellen.

Eine der P-Klassen legt die Wahrscheinlichkeitsverteilung aller Objekte der Wissensbasis fest. Sie ist als Wurzel-P-Klasse bezeichnet (P^*).

Die Knoten einer P-Klasse sind mit den atomaren Konzepten der Wissensbasis assoziiert. Ein Knoten kann den Wert „T“ oder „F“ annehmen, abhängig davon, ob das Konzept, das er modelliert, gilt oder nicht. Zusätzlich sind Knoten für negierte atomare Konzepte eingeführt. Sie hängen nur von ihren nicht negierten Pendants ab. Der Sinn der Einführung dieser Knoten wird dann deutlich, wenn man versucht z.B. die Wahrscheinlichkeit der folgenden logischen Aussage zu ermitteln: $P(C \wedge \neg C)$. Der richtige Wert ist 0 und muß sich aus der entsprechenden Wahrscheinlichkeitsverteilung ergeben.

Darüber hinaus enthält eine P-Klasse für jede Rolle R den Knoten Number(R) zur Modellierung der Anzahl der Rollenfüller. Die Spannweite der Werte dieses Knotens liegt zwischen 0 und einer bestimmten natürlichen Zahl *bound*(R). Damit die Inferenzen in *P*-Classic in endlicher Zeit gemacht werden, ist die Anzahl der Füller einer bestimmten Rolle nach oben begrenzt. Die Wahrscheinlichkeit, mit der Number(R) einen Wert n annimmt, kennzeichnet die Wahrscheinlichkeit, mit der ein Objekt mit n anderen Objekten durch die Rolle R verbunden ist. Das Netz enthält einen weiteren Sonderknoten - PC(R). Er dient zur Bestimmung der Wahrscheinlichkeitsverteilung für die Rollenfüller. Sein Wertebereich umfaßt die Menge aller P-Klassen. PC(R) ist deterministisch, d.h. es darf für jeder Kombination der Werte der Knoten-Vorgänger nur eine P-Klasse mit der Wahrscheinlichkeit 1 ausgewählt werden. Die Einführung des PC(R)-Knoten verleiht der ganzen Struktur einen rekursiven Charakter, wie es aus dem folgenden Beispiel ersichtlich wird.

2.4.5 Beispiel einer P-Klasse

Zur Veranschaulichung der P-Klassen-Abstraktion wird erneut das im Abschnitt 2.3.5 vorgestellte Fragment der TV-Wissensbasis für Sportsedungen aufgegriffen.

In Abbildungen 7 bzw. 8 sind die Wurzel-P-Klasse „sports-broadcasts“ bzw. die P-Klasse „sports-tools“ präsentiert. Die Wurzel-P-Klasse beschreibt die Wahrscheinlichkeitsverteilung aller Objekte der Wissensbasis. Die P-Klasse „sports-tools“ dient zur Spezifikation der Wahrscheinlichkeitsverteilung der Sportgeräte als Rollenfüller der Rolle `has-sports-tool`. Die Liste der verwendeten Abkürzungen ist in der Tabelle 4 dargestellt. Die einzelnen Knoten des Bayes-Netzes sind mit den atomaren Konzepten aus der TV-Wissensbasis assoziiert. Die zugehörigen CPT beinhalten die bedingten Wahrscheinlichkeiten der Gültigkeit von Konzepten in Abhängigkeit von anderen Konzepten. Damit ist die Überlappung von Konzepten modelliert. Zum Beispiel ist in dem Netz kodiert, daß eine TV-Sendung mit der Wahrscheinlichkeit 0.3 über eine „Mannschaftssportart" berichtet, wenn bekannt ist, daß es um eine Sportsendung aber keine „Individualsport"-Sendung handelt.

Die Wahrscheinlichkeit eines Konzepts, das ausschließlich aus Konjunktionen von negierten oder nicht-negierten atomaren Konzepten besteht, ist unmittelbar aus dem Netz ablesbar. Zum Beispiel läßt sich die Wahrscheinlichkeit der Aussage

(`and` *sports-broadcast*
  (`not` *individual-sports-broadcast*)
  team-sports-broadcast)

wie folgt in der Wurzel-P-Klasse ableiten:

$$\begin{aligned}&P(SB=T, ISB=F, TSB=T) = \\ &\quad P(SB=T)\cdot P(ISB=F|SB=T)\cdot P(TSB=T|ISB=F, SB=T) = \\ &\quad 0.1\cdot 0.7\cdot 0.3 = 0.021\end{aligned}$$

Die Berechnung basiert auf der im Abschnitt 2.4.1 erläuterten Kettenregel für das Produkt von zufälligen Ereignissen. In dem betrachteten Beispiel ist die Variablenbelegung komplett spezifiziert. Sind die Werte der Vorgängerknoten nicht bekannt, wird auf die Berechnungsvorschrift für Randwahrscheinlichkeiten zurückgegriffen. Beispielsweise läßt sich die Wahrscheinlichkeit des Konzepts ISB demnach folgendermassen ausrechnen:

$$\begin{aligned}&P(ISB=T) = P(SB=T, ISB=T) + P(SB=F, ISB=T) = \\ &\quad P(SB=T)\cdot P(ISB=T|SB=T) + P(SB=F)\cdot P(ISB=T|SB=F) = \\ &\quad 0.1\cdot 0.3 + 0.9\cdot 0.0 = 0.03\end{aligned}$$

Damit sind die einfachen Inferenzen möglich, wie z.B. die probabilistische Subsumption von Konzepttermen, die aus atomaren und negierten atomaren Konzepten gebildet sind ($P(C|D)$).

Für die Rolle `has-sports-tool` beinhaltet jede P-Klasse zwei Sonderknoten - Number(`has-sports-tool`) und PC(`has-sports-tool`). Der erste definiert die Wahrscheinlichkeitsverteilung der Anzahl der Rollenfüller. Die maximale Anzahl der Sportgeräte ist in dem Beispiel auf 1 beschränkt. Wenn ein Objekt eine Instanz des Konzepts `sports-broadcast` ist, dann steht es in unserer Domäne mit höchstens einem Objekt vom Typ „Sportgerät" in der Beziehung `has-sports-tool`. Eine Instanz des Konzepts `sports-tool` kann dagegen

PC-HST	PC(`has-sports-tool`)
N-HST	Number(`has-sports-tool`)
SB	`sports-broadcast`
ST	`sports-tool`
TSB	`team-sports-broadcast`
ISB	`individual-sports-broadcast`
NISB	¬ `individual-sports-broadcast`
B	`basketball`
F	`football`
TR	`tennis-racket`

Tabelle 4: Verwendete Abkürzungen in der Wissensbasis für TV-Sportsendungen.

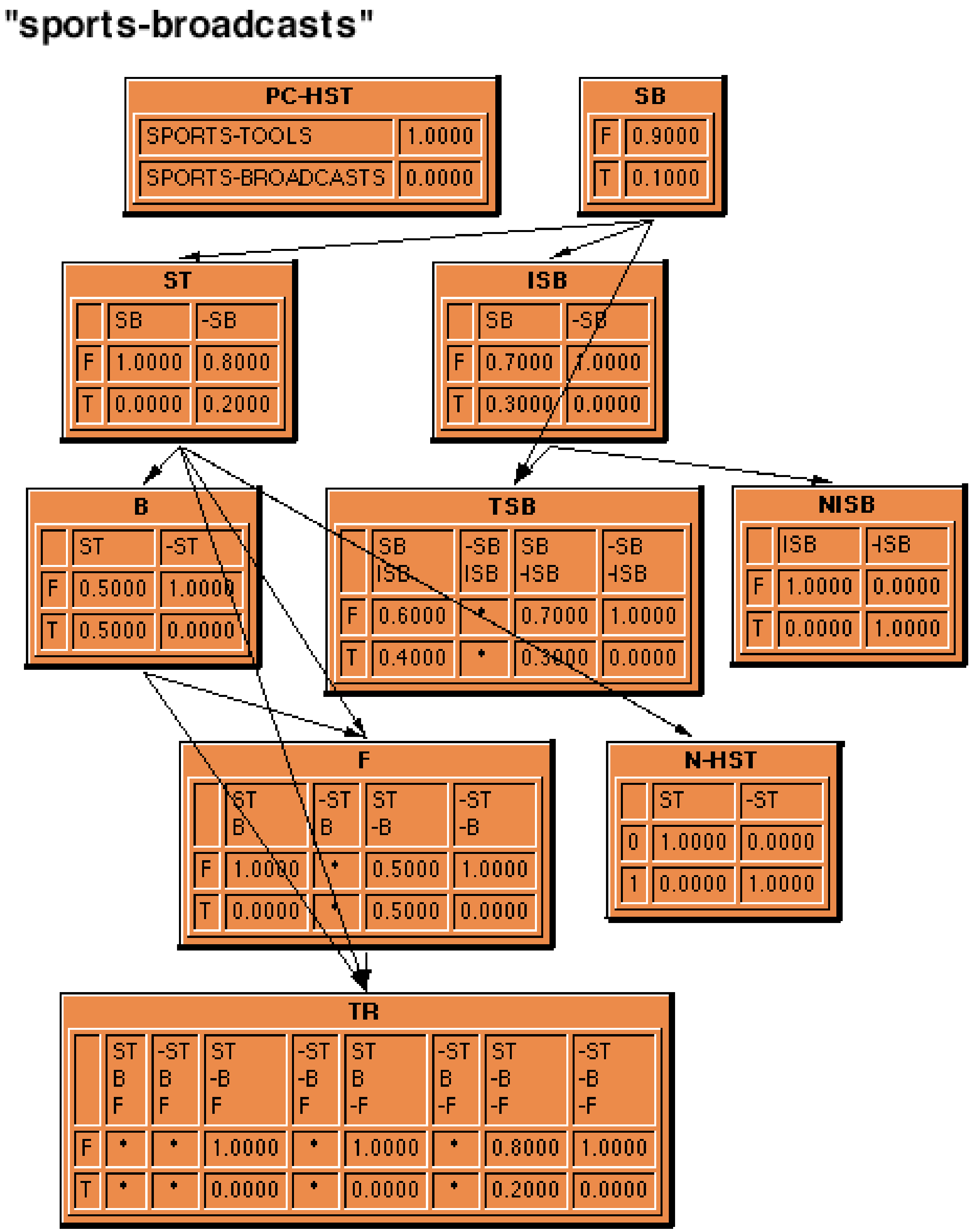

Abbildung 7: P-Klasse „sports-broadcasts“.

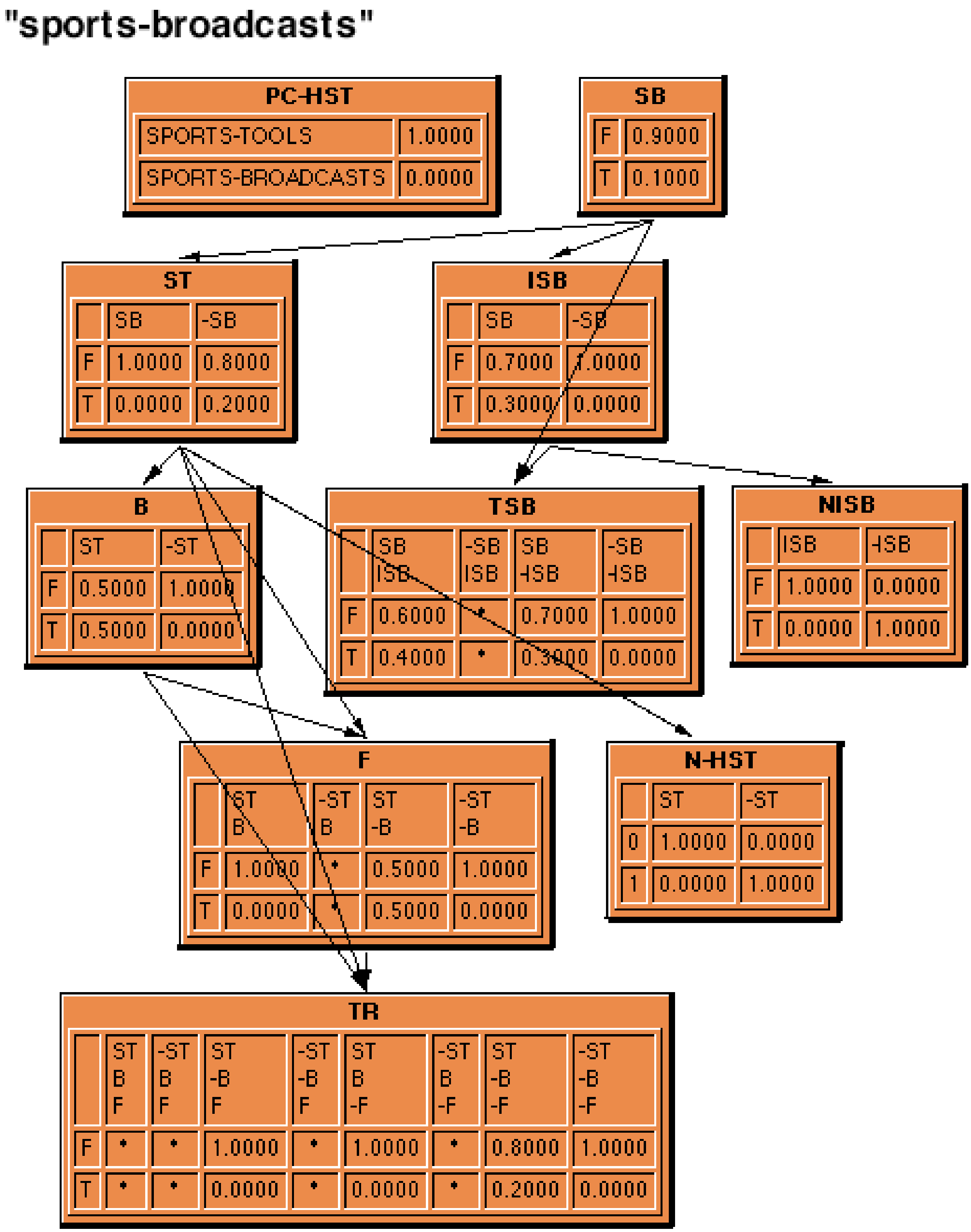

Abbildung 8: P-Klasse „sports-tools“.

unmöglich mit anderen Objekten vom Typ „Sportgerät“ durch die Rolle `has-sports-tool` verbunden sein. Diese Tatsachen sind im Netz wie folgt abgebildet (siehe den Knoten N-HST): Die P-Klasse `sports-broadcasts` definiert die Wahrscheinlichkeit, daß ein Objekt einen Füller der Rolle `has-sports-tool` besitzt, als gleich 1, vorausgesetzt es ist bekannt, daß das Objekt keine Instanz des Konzepts `sports-tool` ist. In der P-Klasse `sports-tools` ist diese Wahrscheinlichkeit dagegen immer 0, da es hier um die Objekte vom Typ „Sportgerät“ handelt und Sportgeräte stehen nicht über die Rolle `has-sports-tool` zu anderen Sportgeräten in Beziehung.

Der zweite Sonderknoten PC(`has-sports-tool`) dient zum Verweis auf die richtige Wahrscheinlichkeitsverteilung der Eigenschaften der Rollenfüller.

Zur Veranschaulichung läßt sich als Beispiel die Wahrscheinlichkeit der Beschreibung $TSB \sqcap (\geq\ 1\ HST) \sqcap (\leq\ 1\ HST) \sqcap \forall\ HST.TR$ wie folgt rekursiv berechnen. Der Knoten `team-sports-broadcast` erhält den Wert „T“. Die Anzahl der Rollenfüller für die Rolle `has-sports-tool` ist 1, deshalb wird der Knoten Number(`has-sports-tool`) mit 1 initiiert. Die Rollenfüller werden anhand der P-Klasse „sports-tools“ selektiert, also bekommt der Knoten PC(`has-sports-tool`) den Wert „sports-tools“ und der Knoten `tennis-racket` hat den Wert „T“. Die resultierende Wahrscheinlichkeit (0.00624) kann dann mit bereits erwähnten Inferenzmechanismen in Bayes-Netzen ermittelt werden.

Bevor ein allgemeines Verfahren zur Berechnung der Wahrscheinlichkeiten von Konzepten in kanonischer Form vorgestellt wird, muß noch die Semantik der probabilistischen Komponente von *P*-Classic beschrieben werden.

2.4.6 Semantik der probabilistischer Erweiterung

Die probabilistische *P*-Classic-Komponente ergänzt die in 2.3.2 eingeführte Interpretation der T-Box $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ um die Interpretation von P-Klassen. Eine Interpretation $\mathcal{I}^p$ der probabilistischen Komponente von *P*-Classic ist eine Menge der Wahrscheinlichkeitsverteilungen $\mu^{\mathcal{I}}_{P_j}$ der Domäne $\Delta^{\mathcal{I}}$. Jede dieser Wahrscheinlichkeitsverteilungen ist mit einer P-Klasse P_j assoziiert.

Eine Wahrscheinlichkeitsverteilung $\mu^{\mathcal{I}}_{P_j}$ der Domäne $\Delta^{\mathcal{I}}$ kann nur dann als *konsistent* bzgl. der P-Klasse P_j betrachtet werden, wenn die folgenden Bedingungen zutreffen:

1. Für jedes Konzept C in der kanonischen Form ist die Wahrscheinlichkeit von $C^{\mathcal{I}}$ identisch mit der Wahrscheinlichkeit, die für C im Bayes-Netz der P-Klasse P_j definiert ist:
$$\mu_{P_j}(C^{\mathcal{I}}) = BN_{P_j}(C)$$
2. Für jedes Konzept C und jede Rolle R sei a eine Instanz von C ($a^{\mathcal{I}} \in C^{\mathcal{I}}$) und b ein Rollenfüller der Rolle R bzgl. a ($(a^{\mathcal{I}}, b^{\mathcal{I}}) \in R^{\mathcal{I}}$) im Modell $\mathcal{I}$. Die Wahl der Rollenfüller b stellt ein zufälliges Ereignis dar, welches durch die Wahrscheinlichkeitsverteilung μ_{P_i} beschrieben ist. Dann muß μ_{P_i} bzgl. der P-Klasse P_i konsistent sein. P_i ist hier der Wert des deterministischen Knoten PC(R) in der P-Klasse P_j unter der Bedingung, daß C gilt:
$$P_i = BN_{P_j}(PC(R)|C)$$
3. Die Rollenfüller einer Rolle R werden unabhängig voneinander und von Füllern anderer Rollen gewählt. Die Wahrscheinlichkeitsverteilung der Rollenfüller hängt mit anderen Worten nicht von den Eigenschaften anderer Rollenfüller ab, sondern wird bzgl. der

Rolle R eindeutig festgelegt. Aus dieser Annahme folgt eine einfache Berechnungsregel für die Wahrscheinlichkeit mehrerer Rollenfüller: Ist die Wahrscheinlichkeit, daß ein Objekt a einen Rollenfüller hat, gleich ρ, so ist die Wahrscheinlichkeit, daß es n Rollenfüller hat, gleich ρ^n.

Desweiteren wird angenommen, daß die Wahrscheinlichkeitsverteilung der Rollenfüller für ein Objekt auch nicht von den Eigenschaften des Objektes abhängt. Zum Beispiel wird die Wahrscheinlichkeitsverteilung von Füllern der Rolle HST (beschrieben durch die P-Klasse „has-sports-tools") nicht dadurch beeinflußt, ob das Konzept

$$TSB \sqcap (\geq\ 1\, HST) \sqcap (\leq\ 1\, HST) \sqcap \forall\, HST.TR$$

oder das Konzept

$$ISB \sqcap (\geq\ 1\, HST) \sqcap (\leq\ 1\, HST) \sqcap \forall\, HST.TR$$

gilt (siehe Abbildungen 7 bzw. 8)

Ein probabilistisches terminologisches Axiom besitzt die folgende Form:

$$P(C|D) = \rho$$

C und D sind Konzepte, ρ steht für einen Wahrscheinlichkeitswert aus dem Wertebereich [0,1]. Ist $D = \top$, so verkürzt sich die oben aufgeführte Definition zum folgenden Ausdruck:

$$P(C) = \rho$$

Die Interpretation $\mathcal{I}^p$ erfüllt ein probabilistisches terminologisches Axiom $P(C|D) = \rho$ genau dann, wenn der Wahrscheinlichkeitswert ρ aus dem Bayes-Netz für die Wurzel-P-Klasse P^* abgeleitet werden kann: $\mu^{\mathcal{I}}_{P^*}(C|D) = \rho$.

Gelten die oben aufgezählten Konsistenzbedingungen, dann hat jedes Konzept C in der *P*-Classic-Wissensbasis eine eindeutige Wahrscheinlichkeit $\rho \in [0,1]$, und das probabilistische Axiom $P(C) = \rho$ ist erfüllt in Bezug auf die *P*-Classic-Wissensbasis. Den Beweis findet man in [7].

Wie bereits angedeutet wurde, betreffen die bisherigen Definitionen eine probabilistische Erweiterung der T-Box. Die A-Box kann ebenfalls entsprechend modifiziert werden, um die Deklaration von probabilistischen Behauptungen (assertions) der folgenden Form zu ermöglichen: $P(a \in C) = \rho$, a ist ein Individuum und ρ ein Wahrscheinlichkeitswert. In *P*-Classic wurde die Behandlung der probabilistischen Zusicherungen ausgelassen und auch im Rahmen dieser Arbeit nicht weiter verfolgt.

2.4.7 Inferenzverfahren

Im Abschnitt 2.4.5 wurden einige Beispiele für probabilistische Inferenzen mit P-Klassen gezeigt. In diesem Teil der Arbeit soll ein allgemeines Verfahren zur Berechnung der Wahrscheinlichkeiten von Konzepttermen in *P*-Classic erläutert und anhand weiterer Beispiele illustriert werden.

Annahmen Es wird vorausgesetzt, daß Konzepte sich in der kanonischen Form befinden und die Tiefe der Konzeptterme eingeschränkt ist (vgl. 2.3.4). Die naheliegende Methode, die Wahrscheinlichkeit eines Konzeptausdrucks

$$C = \alpha \sqcap \beta_{R_1} \sqcap \ldots \sqcap \beta_{R_k} \mid \beta_{R_i} = (\geq\ l_{R_i}\ R_i) \sqcap (\leq\ m_{R_i}\ R_i) \sqcap \forall\ R_i.C_{R_i}, i \in [1, k]$$

zu berechnen, wäre als erstes die Wahrscheinlichkeit des Teilausdrucks α zu bestimmen, dann rekursiv die Wahrscheinlichkeiten der β-Teile mit bekannten Schlußfolgerungsregeln der Bayes-Netze zu berechnen. Die Laufzeit dieser Vorgehensweise ist aber exponenziell zu der Tiefe des Konzepts C. Das Ziel besteht darin, ein effizientes Verfahren zu entwickeln. Der in [7] vorgeschlagene Algorithmus basiert auf zwei Beobachtungen. Zum einen ist es sinnvoll, die Berechnung mit den Teilausdrücken der Tiefe 0 anzufangen. Es werden dabei die Wahrscheinlichkeiten der atomaren Konzepte C_{R_i} ermittelt und für die späteren Berechnungen aufgehoben. Die nächsten Berechnungsschritte verlaufen „bottom up" bzgl. der Konzepttiefe, und in jedem Schritt werden die ermittelten Wahrscheinlichkeiten für die Verwendung im nächsten Schritt gespeichert. Die zweite Beobachtung besteht darin, daß die Wahrscheinlichkeit eines β_{R_i}-Teils nur von der Anzahl und der P-Klasse der Rollenfüller der Rolle R_i abhängt. Der Knoten β_{R_i} ist mit anderen Worten „d-separiert" vom Rest des Netzes durch die Knotenmenge $\{$PC(R_i), Number(R_i)$\}$ (s. Definition im Abschnitt 2.4.3). Deshalb kann man den β_{R_i}-Knoten in einer P-Klasse direkt unter die entsprechenden Knoten PC(R_i) und Number(R_i)-Knoten hängen. Die bedingte Wahrscheinlichkeitsverteilung des neuen β_{R_i}-Knotens läßt sich für jede Kombination der Knoten-Vorgänger ausrechnen und wird in die CPT von β_{R_i} eingetragen.

Beschreibung Der vollständige Algorithmus **Compute Probability** zur Berechnung der Wahrscheinlichkeit eines Konzepts C bzgl. einer *P*-Classic-Wissensbasis sieht wie folgt aus:

Es kann leicht gezeigt werden, daß die Komplexität des Algorithmus linear zu der Tiefe des Konzepts C und quadratisch zu der Anzahl der P-Klassen ist. Sollten die Bayes-Netze der P-Klassen so strukturiert sein, daß Inferenzen in polynomaler Zeit möglich sind, so bleibt die Laufzeit des Algorithmus auch polynomial bzgl. der Größe der Wissensbasis.

Beispiel Als Beispiel wird die Wahrscheinlichkeit des Konzepts

$$C = TSB \sqcap (\geq\ 1\ HST) \sqcap (\leq\ 1\ HST) \sqcap \forall\ HST.B$$

ermittelt. Für die Berechnungen werden die im Abschnitt 2.4.5 eingeführten P-Klassen für die TV-Wissensbasis herangezogen. Die Tiefe des Konzepts C ist 1. Der Teilausdruck B hat die Tiefe 0. Im ersten Schritt wird die Wahrscheinlichkeit von B für jede der P-Klassen „sports-broadcasts" bzw. „sports-tools" mit Hilfe der Inferenzregeln für Bayes-Netze ermittelt und gespeichert:

$$P^0_{\text{sports-broadcasts}}(B) = 0.09$$

$$P^0_{\text{sports-tools}}(B) = 0.27$$

Als nächstes muß die Wahrscheinlichkeitsverteilung des β-Teilausdrucks $(\geq\ 1\ HST) \sqcap (\leq\ 1\ HST) \sqcap \forall\ HST.B$ bestimmt werden. Der Algorithmus sieht vor, daß zu diesem Zweck ein neuer Knoten in der Wurzel-P-Klasse „sports-broadcasts" eingesetzt wird, der nur von

Knoten Number(HST) und PC(HST) bedingt abhängt. Für jede Kombination der Werte von Number(HST) und PC(HST) sollen nun die bedingten Wahrscheinlichkeiten von β gefunden und in die CPT des β-Knotens eingetragen werden. Die CPT von β zeigt Abbildung 10.

Da die Anzahl der Rollenfüller für die Rolle HST in β auf 1 eingeschränkt ist, wird die Wahrscheinlichkeit, daß es mehr oder weniger als 1 Rollenfüller gibt, mit 0 bewertet. Die Wahrscheinlichkeit, daß es genau einen Rollenfüller gibt, wurde in dem ersten Schritt für jede P-Klasse bereits ermittelt (0.09 bzw. 0.27). Vorausgesetzt die Annahme, daß die Rollenfüller voneinander unabhängig sind, läßt sich die Wahrscheinlichkeit, daß es 2, 3, usw. Füller gibt, im allgemeinen als ein Produkt der Wahrscheinlichkeiten jedes Füllers berechnen. Demnach würde die Wahrscheinlichkeit der Existenz von 2 Rollenfüllern in der P-Klasse „sports-broadcasts" in dem betrachteten Beispiel ohne die Anzahlrestriktion folgendermaßen ermittelt:

$$P^0_{\text{sports-broadcasts}}(B))^2 = 0.09^2 = 0.0081$$

Abbildung 11 zeigt die Wurzel-P-Klasse „sports-broadcasts" mit dem eingefügten β-Knoten. Jetzt sind alle Informationen für die Berechnung der gemeinsamen Wahrscheinlichkeitsverteilung der α- und β-Teilausdrücke vorhanden. Die Wahrscheinlichkeit des Konzepts C läßt sich aus der P^*-Klasse „sports-broadcasts" mit üblichen Bayes-Netz-Inferenzmethoden ableiten:

$$P(C) = P(TSB = T, ((\geq\ 1\,HST) \sqcap (\leq\ 1\,HST) \sqcap \forall\,HST.B) = T) = 0.00891$$

2.5 P-Classic-Probleme und Diskussion

Die Annahme der Unabhängigkeit der Rollenfüller stellt eine der entscheidenden Voraussetzung für die Traktabilität der P-Classic-Beschreibungslogik dar. Es wurden diverse Vorschläge gemacht, wie diese Einschränkung teilweise aufgehoben werden kann [7]. Eine mögliche Lösung ist z.B. die Einführung von sogenannten „versteckten" (hidden) Konzepten (ähnlich zu den „hidden" Variablen in Bayes-Netzen), die zwar nicht in der Terminologie der Beschreibungslogik enthalten sind, aber in einer P-Klasse die gleiche Rolle wie die anderen atomaren Konzepte spielen. Mit Hilfe von versteckten Konzepten können die Abhängigkeiten zwischen den Rollenfüllern ausgedrückt werden. Die Wahl einer richtigen P-Klasse für die Rollenfüller hängt von den versteckten Konzepten ab. Zusätzlich müssen neue P-Klassen eingeführt werden, um die Abhängigkeit der Rollenfüller vom versteckten Konzept probabilistisch zu spezifizieren. Sollten in einer Wissensbasis viele solche Abhängigkeiten modelliert werden, explodiert die Anzahl der P-Klassen. Um diesen unerwünschten Effekt zu vermeiden, wurde in [7] eine andere Alternative in Erwägung gezogen, die sich momentan noch in einem Stadium der Untersuchung befindet. Die Idee besteht darin, einen Rollenfüller als „eine Funktion" zu betrachten, deren Parameter die Eigenschaften des mit dem Rollenfüller durch eine Rolle verbundenen Objektes sind. Auf diese Weise können die Eigenschaften eines Objektes die Wahrscheinlichkeitsverteilung der Rollenfüller direkt beeinflussen.

Die vorausgesetzte Unabhängigkeit der Rollenfüller führt dazu, daß auch der „fills"-Operator als Problemfall angesehen werden muß. Als erstes soll klargestellt werden, daß die semantische Definition dieses Konstruktes in P-Classic(vgl. 2.4.6) aus der Semantik des vollständigen Classic-Systems übernommen wurde. Das vollständige Classic enthält Konzeptkonstrukte, in denen die Individuen vorkommen (z.B. „same-as", „fills", „one-of").

Es wurde erst später gezeigt, daß die Inferenzen im vollständigen CLASSIC-System nicht traktabel sind. Sie Semantik von CLASSIC wurde entsprechend modifiziert ([3]). Laut dieser neuen Semantik werden die Individuen in Konzepttermen auf die disjunkten Mengen von Domäneobjekten abgebildet. Für jedes Individuum I wird ein atomares Konzept A_I deklariert und es wird durch zusätzliche Axiome in der T-Box zugesichert, daß alle A_I-Konzepte untereinander disjunkt sind. Die A-Box beinhaltet für jedes Individuum I ein assertorisches Konzeptaxiom, welches I dem atomaren Konzept A_I zuweist. Die Semantik des Operators „fills“ sieht in dem modifizierten CLASSIC wie folgt aus:

$$(\text{fills } QI) : \{x \in \Delta^{\mathcal{I}} | \exists y \in \Delta^{\mathcal{I}} : (x, y) \in R^{\mathcal{I}} \wedge y \in {A_I}^{\mathcal{I}}\}$$

In der terminologischen Komponente von *P*-CLASSIC wurde die alte Semantik des „fills“ -Operators erstmal beibehalten. Als ernstes Problem stellte sich allerdings heraus, daß die Verwendung des „fills“-Konstruktes in der probabilistischen Komponente des *P*-CLASSIC-Wissenspräsentationssystems die Annahme der Unabhängigkeit der Rollenfüller verletzten würde. Als Beweis wird z.B. die folgende Aussage betrachtet:

$$C = (\mathsf{fills}\, R\, I_1) \sqcap (\mathsf{fills}\, R\, I_2)$$

C impliziert, daß I_1 und I_2 gleich und deshalb nicht unabhängig sind. Aufgrund dieser Beobachtung sind die „fills“-Knoten in einer P-Klasse nicht zulässig.

Andere Individuen betreffende Konstrukte von CLASSIC („same-as“ und „one-of“) sind aus dem gleichen Grund in die Semantik von *P*-CLASSIC nur schwer zu integrieren. Zum Beispiel postuliert der „same-as“ Operator die Gleichheit der Attributketten. So soll die Aussage (**same-as** $(q_1 q_2 \ldots q_n)$ $((r_1 r_2 \ldots r_n))$ im vollständigen CLASSIC eine Menge von Individuen beschreiben, die sowohl unter dem ersten $(q_1 q_2 \ldots q_n)$ als auch unter dem zweiten $(r_1 r_2 \ldots r_n)$ Attributpfad erreicht werden können. Die Äquivalenz der Individuen, die durch das Fortschreiten entlang der verschiedenen Attributketten erreicht werden können, widerspricht wieder der Grundannahme der Unabhängigkeit von Rollenfüllern in *P*-CLASSIC.

Als nächstes soll noch die Behandlung der Negation in *P*-CLASSIC diskutiert werden. In der Semantik von *P*-CLASSIC ist die Negation von atomaren Konzepten erlaubt. Eine P-Klasse besteht u.a. aus Knoten zur Darstellung von atomaren Konzepten einer Wissensbasis. Ein Knoten für ein Konzept C bekommt einen „True“-Wert, wenn ein Objekt eine Instanz von C ist, und einen „False“-Wert, wenn es keine Instanz von C ist. Die letzte Konstellation wurde in [7] als Negation von C aufgefaßt. Bei dieser Interpretation treten allerdings Schwierigkeiten auf. Angenommen, es wird die Wahrscheinlichkeit der Aussage $C \sqcap \neg C$ berechnet. Entgegen der Erwartung ist das Ergebnis ein Wahrscheinlichkeitswert, der größer als 0 ist. Die Begründung dafür liegt darin, daß Ereignisse, in denen ein Knoten C den einen oder den anderen Wert annimmt, voneinander unabhängig sind. Die Disjunktheit von negierten und nicht negierten atomaren Konzepten mit gleichen Namen muß deshalb gesondert modelliert werden. Aufgrund dieser Überlegung wurde für jeden Knoten zur Beschreibung eines atomaren Konzepts jeweils ein Knoten für das negierte atomare Konzept eingeführt und mit der entsprechenden Wahrscheinlichkeitsverteilung versehen. Gilt C, so muß die bedingte Wahrscheinlichkeit von $\neg C$ gleich 0 sein.

Anschließend muß noch erwähnt werden, daß bestimmte Implikationen in *P*-CLASSIC noch nicht möglich sind. Wenn ein Konzept C die folgende Form hat: $C = (\forall\, R.D)$ und die Wahrscheinlichkeit des Konzepts D gleich 0 ist, dann muß daraus folgen, daß die Rolle R

keine Füller hat. Das heißt, es gilt die folgende Äquivalenz, die in *P*-Classic noch nicht richtig erkannt wird:

$$((\forall\ R.D) \mid P(D) = 0) \equiv ((\geq\ 0\ R) \sqcap (\leq\ 0\ R))$$

Es ist darauf hinzuweisen, daß bis jetzt die Rede nur von einer probabilistischen Erweiterung der T-Box war. Prinzipiell spricht auch nichts gegen die Definition einer probabilistischen A-Box. Die probabilistischen Behauptungen in der A-Box sollen das unsichere Wissen über die Individuen der Anwendungsdomäne ausdrücken. Zum Beispiel würde man in einer probabilistischen A-Box für eine konkrete „Tennis-Davis-Cup"-Sendung (a) die Wahrscheinlichkeit angeben, mit der sie eine Instanz des Konzepts `individual-sports-broadcast` (C) ist: $P(a \in C) = 0.8$. Wie schon [6] verdeutlichte, ist die Natur dieser subjektiven Wahrscheinlichkeit von Grund auf verschieden von der Wahrscheinlichkeit, mit der sich Konzepte subsumieren. Der Wahrscheinlichkeitsverteilung der Konzeptsubsumption liegen von vornherein bekannte statistische Informationen zugrunde, die als Ergebnis von Beobachtungen einer Reihe konkreter Anwendungsobjekte und ihrer Zugehörigkeit einem oder anderem Konzept entstanden sind. In *P*-Classic wurden alle mit einer probabilistischen A-Box zusammenhängenen Fragen, so auch die Problematik der induktiven Schlußfolgerungen in einer probabilistischen Beschreibungslogik, erstmal nicht betrachtet. Sie stellen einen interessanten Ansatzpunkt für weitere Forschungsarbeiten dar.

Zu den Zielen der Weiterentwicklung von *P*-Classic gehören u.a. die Optimierung des Inferenzverfahrens. Im ersten Prototyp wurde nur ansatzweise die Zwischenspeicherung (Caching) der berechneten Wahrscheinlichkeiten implementiert. Es können andere ausgeklügelten Techniken überlegt werden, um die Laufzeit der Inferenzen in *P*-Classic zu verringern. Auf jeden Fall sollen die probabilistischen Inferenzen in den Bayes-Netzen so optimal wie möglich berechnet werden. In diesem Bereich gibt es ein breites Spektrum wissenschaftlicher Untersuchungen.

Die terminologische Komponente von *P*-Classic stellt in diesem ersten Stadium der Entwicklung und Implementation eine noch ziemlich eingeschränkte Variante der Beschreibungslogik Classic dar (mit Ausnahme des Negationsoperators für atomare Konzepte). Eine ausdrucksmächtigere Terminologie ist gewünscht.

Compute Probability (C, KB)

Initialization: C ist ein Konzept in kanonischer Form,

  KB ist eine P-Classic-Wissensbasis,

  $R_1, \ldots, R_m$ sind Rollen, P^* ist die Wurzel-P-Klasse

  for d = 0 bis zu der Tiefe des Konzepts C: depth(C)) **do**

    for jede P-Klasse P in KB **do**

      for jeden Teilausdruck $C = \alpha \sqcap \beta_{R_1} \sqcap \ldots \sqcap \beta_{R_m}$ der Tiefe d **do**

        for jeden Teilausdruck $\beta_{R_j} = (\geq l_j R_j) \sqcap (\leq u_j R_j) \sqcap (\forall R_j.D_j)$ in C, $1 \geq j \leq m$ **do**

          for h=0 bis zur maximalen Anzahl der Rollenfüller für die Rolle R_j:

          bound (R_j) **do**

            for jede P-Klasse P **do**

              if $u_j \geq h \geq l_j$ **then**

                $CPT(h, P, T) = P(D_j)^h$

              else

                $CPT(h, P, T) = 0$

          füge den Knoten β_{R_j} in die P-Klasse P ein

        berechne und speichere die Wahrscheinlichkeit von C in P

  gib die Wahrscheinlichkeit von C in P^* zurück

Abbildung 9: Der P-Classic-Inferenzalgorithmus **Compute Probability**.

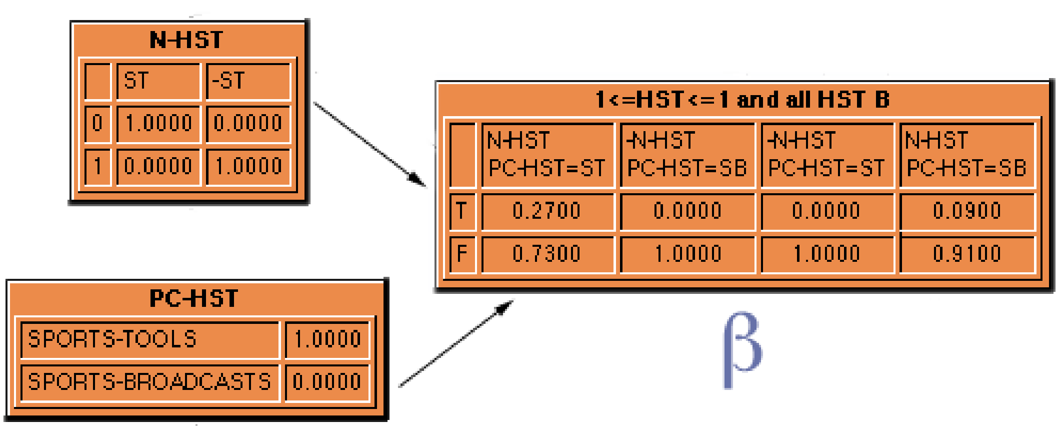

Abbildung 10: Die Wahrscheinlichkeitsverteilung des β-Knoten $(\geq\ 1\, HST) \sqcap (\leq\ 1\, HST) \sqcap \forall\, HST.B$.

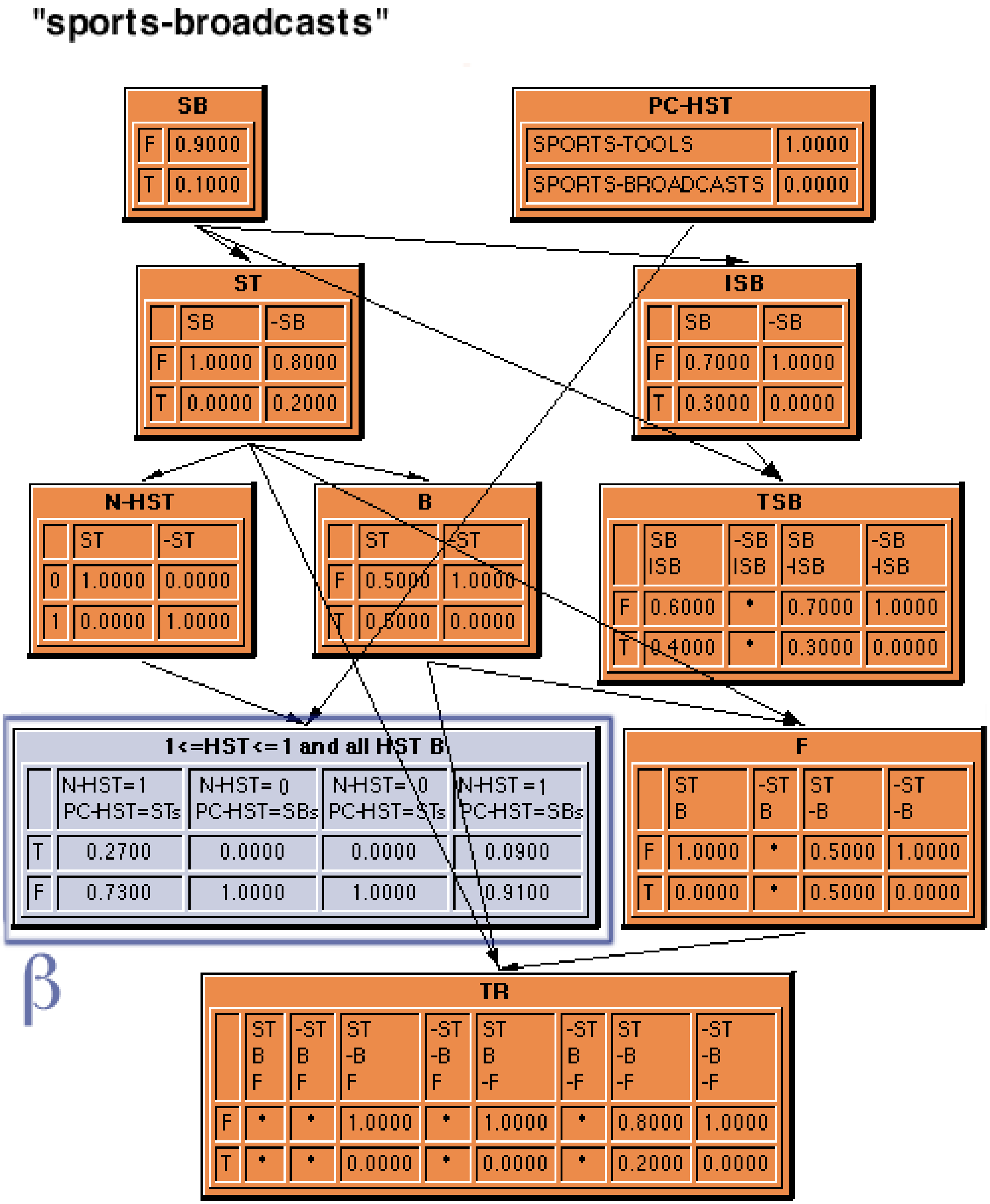

Abbildung 11: P-Klasse „sports-broadcasts“ mit dem β-Knoten.

3 PLCS

Die Aufgaben der Schlußfolgerungsmechanismen in Beschreibungslogiken umfassen einerseits die Konsistenzübeprüfung der T- und A-Box Axiome und andererseits das Ableiten von implizitem Wissen aus den in der Wissensbasis vorhandenen expliziten Deklarationen. Im Abschnitt 2.3.3 wurden die wichtigsten von Inferenzdiensten aufgeführt, dazu gehört z.B. das Testen der Konzeptsubsumption. Dieses Kapitel widmet sich dem beschreibungslogischen „Least Common Subsumer"-Operator und dessen probabilistischen Variante. Dem LCS-Operator ist insbesondere bei der beispielorientierten Informationsrecherche eine wichtige Rolle zugeteilt. Wie schon allein aus der Bezeichnung geschlossen werden kann, handelt es sich um einen Konstruktionsoperator, dessen Zweck besteht darin, für zwei oder mehrere Konzepte die speziellste Konzeptbeschreibung zu generieren, welche die Konzepte subsumiert. Die Grundlagen des LCS-Operators werden im folgenden insofern umrissen, als sie für die Motivation einer probabilistischen Weiterentwicklung des LCS-Operators von Bedeutung sind. Das im vorigen Kapitel vorgestellte System *P*-Classic ermöglichte die Entwicklung des *P*LCS-Operators, auf dessen Eigenschaften und Berechnungsmethoden in diesem Teil der Arbeit eingegangen wird. Es wird an Beispielen gezeigt, daß unter Anwendung des *P*LCS-Operators statt des LCS-Operators in einem beispielorientierten Informationsrecherche-Kontext in vielen Fällen deutlich bessere Ergebnisse erzielt werden können.

3.1 Die Grundidee des LCS-Operators

Die in diesem Abschnitt verwendeten Definitionen basieren größtenteils auf den Arbeiten von W. Cohen und A. Borgida ([4]) bzw. C. Prien ([18]).

Der LCS-Operator erzeugt für zwei oder mehrere Konzepte einen neuen Konzeptterm, der die Gemeinsamkeiten der Konzepte beschreibt. Eine formale Definition folgt. Das Hinzufügen der LCS-Methode zu den üblichen beschreibungslogischen Inferenzverfahren wurde durch eine Reihe nutzvoller Anwendungen motiviert. Bei induktiven Lernprozessen stellt z.B. die Suche nach einem speziellsten gemeinsamen Konzept eine gebräuchliche Vorgehensweise dar, um die Lernbeispiele zu generalisieren. Die Inferenzen in den Wissensbasen mit disjunkten Konzepten sind oftmals nicht traktabel. Eine der Lösung dieses Problems ist die sogenannte logische „Vivifikation" (Lebendigmachung) der Wissensbasis. Die Disjunktionen werden dabei durch das LCS-Konzept der Disjunkte ersetzt und dadurch die Traktabilität der auf diese Weise approximierten Wissensbasis gewährleistet. Wir verwenden die LCS-Methode, wie bereits erwähnt, für eine beispielorientierte Informationsrecherche.

3.1.1 Definition des LCS

Ein Konzept C wird als „Least Common Subsumer" der Konzepte D_1 und D_2 bezeichnet, wenn folgendes gilt:

1. C subsumiert sowohl D_1 als auch D_2;
2. es gibt kein anderes Konzept C', welches D_1 und D_2 subsumiert, selbst aber von C echt subsumiert wird.

Formal ausgedrückt:

$$\neg\exists C' : D_1 \preceq C, D_2 \preceq C, C' \preceq C \wedge \neg(C \preceq C')$$

Im Kontext einer beliebigen Beschreibungslogik $\mathcal{L}$ können durchaus mehrere semantisch ungleiche LCS-Konzepte existieren. Die Aufgabe besteht darin, für jede semantische Äquivalenzklasse nur einen LCS-Kandidaten auszuwählen. Diese Aufgabe übernimmt der LCS-Operator, der wie folgt definiert ist:

$\mathcal{L}$ ist eine Beschreibungslogik, C_1 und C_2 sind Konzepte in $\mathcal{L}$. Der LCS-Operator von C_1 und C_2 ($\mathrm{LCS}(C_1, C_2)$) ist eine Konzeptmenge $lcs = \{D_1, D_2, \ldots, D_n\}$, so daß die folgenden Bedingungen erfüllt sind:

1. jedes Konzept $D_i \in lcs$ subsumiert C_1 und C_2;
2. alle Konzepte der Menge lcs sind untereinander semantisch unterschiedlich: $D_i, D_j \in lcs, i \neq j : D_i \not\equiv D_j$
3. für jede Äquivalenzklasse enthält lcs nur einen Repräsentanten, d.h. wenn ein LCS-Konzept D von C_1 und C_2 existiert, dann ist in der Menge lcs ein Konzept D_i vorhanden, so daß $D \equiv D_i$ gilt.

Wenn die Beschreibungslogik $\mathcal{L}$ einen Schnittmengenoperator zur Verfügung stellt, dann kann bewiesen werden, daß die Menge lcs aus genau einem Element besteht. CLASSIC und entsprechend P-CLASSIC haben den Operator „and“ und deshalb ist das LCS-Konzept eindeutig.

Weitere Eigenschaften des LCS-Operators sind Kommutativität und Assoziativität. Daraus folgt, daß LCS für mehr als zwei Argumente durch folgende Formel ermittelt werden kann:

$$\mathrm{LCS}(C_1, C_2, \ldots, C_n) = \mathrm{LCS}(\mathrm{LCS}(\ldots \mathrm{LCS}(C_1, C_2) \ldots), C_n)$$

Die Berechnung des LCS-Operators scheint mit der Berechnung der Subsumptionsrelation etwas Gemeinsames zu haben. In [4] wurde an Beispielen jedoch gezeigt, daß dies generell nicht immer zutrifft. Zum Beispiel gibt es Beschreibungslogiken, für die die Schlüsse über die Subsumptionsrelation in polynomialer Zeit gemacht werden können, während der LCS-Operator nicht in polynomialer Zeit berechnet werden kann. Dennoch gibt es einen Spezialfall, in dem der LCS-Operator mit der Subsumption korreliert. Es handelt sich dabei um die strukturelle Subsumption, wie sie neben anderen Beschreibungslogiken auch in CLASSIC realisiert ist.

Die strukturelle Subsumption In einem strukturellen Subsumptionsverfahren wird die Subsumption von Beschreibungen durch den voneinander unabhängigen Vergleich von Teilbeschreibungen getestet, die mit den gleichen Operatoren konstruiert sind. Die Voraussetzung dafür ist, daß sich die Konzepte in Normalform befinden. Die Normalisierung von Konzepten bewirkt u.a., daß impizite Fakten expliziert und Inkonsistenzen erkannt werden. Die definierten Konzepte werden dabei entfaltet (z.B. FB durch den Term auf der rechten Seite $TSB \sqcap (\geq\ 1\ HST) \sqcap (\leq\ 1\ HST) \sqcap \forall\ HST.F$ ersetzt). Die Konzepte A, für die in der T-Box ein Axiom der Form $A \sqsubseteq C$ existiert, werden durch $A \sqcap C$ substituiert. Mittels Normalisierung wird die gegenseitige Unabhängigkeit der einzelnen Konstrukte der Konzeptbeschreibungen sichergestellt. Zum Beispiel wird erkannt, daß die Beschreibung $TSB \sqcap (\geq 2\, HST) \sqcap (\leq 1\, HST) \sqcap \forall HST.TR$ inkonsistent ist und deshalb von jeder anderen Beschreibung subsumiert wird.

Operator k	Partielle Ordnungsrelation $\leq_k$ (`Subsumiert?`$((k\ \alpha),(k\ \beta))$)
and	$(C_1,\ldots,C_n) \leq_{\text{and}} (D_1,\ldots,D_m) = \begin{cases} \text{true} & \text{wenn } \forall\ C_i \mid \exists\ D_j : C_i \Rightarrow D_j \\ \text{false} & \text{sonst} \end{cases}$
all	$(R_1,C_1) \leq_{\text{all}} (R_2,C_2) = \begin{cases} \text{true} & \text{wenn } R_1 = R_2 \text{ und } C_1 \Rightarrow C_2 \\ \text{false} & \text{sonst} \end{cases}$
at-least	$(l_1,R_1) \leq_{\text{at-least}} (l_2,R_2) = \begin{cases} \text{true} & \text{wenn } R_1 = R_2 \text{ und } l_1 \geq l_2 \\ \text{false} & \text{sonst} \end{cases}$
at-most	$(u_1,R_1) \leq_{\text{at-most}} (u_2,R_2) = \begin{cases} \text{true} & \text{wenn } R_1 = R_2 \text{ und } u_1 \leq u_2 \\ \text{false} & \text{sonst} \end{cases}$
atomic concept	$(C_1) \leq_{\text{atomic}} (C_2) = \begin{cases} \text{true} & \text{wenn } C_1 = C_2 \text{ und } C_i \sqsubseteq \top \\ \text{false} & \text{sonst} \end{cases}$

Tabelle 5: Strukturelle Subsumption in CLASSIC.

Die Subsumptionsrelation von normalisierten Konzepten hängt nur von der Struktur der Konzepte ab und kann mit relativ einfachen Algorithmen ermittelt werden. Hierfür wird ein Operator `Subsumiert?` von zwei Argumenten definiert. Der Operator `Subsumiert?`$((k_1\ \alpha),(k_2\ \beta))$ gibt „false" zurück, wenn $k_1 \neq k_2$ gilt. k_1, k_2 sind termbildende Operatoren. Beispielsweise liefert `Subsumiert?`$((\geq\ 1\, HST),\ (\leq\ 1\, HST))$ „false", weil die Operatoren ungleich sind. Sind die Operatoren gleich ($k_1 = k_2 = k$), dann hängt die Subsumptionsrelation von α bzw. β und vom angewendeten Operator k ab. Diese Beziehung kann mit dem Begriff „partielle Ordnungsrelation" ausgedrückt werden. Die partielle Ordnungsrelation bzgl. des Operators k wird als $\leq_k$ denotiert. Die Terme $(k\ \alpha)$ und $(k\ \beta)$ befinden sich in der strukturellen Subsumptionsrelation `Subsumiert?`$((k\ \alpha),(k\ \beta))$ genau dann, wenn die partielle Ordnungsrelation zwischen α und β in Bezug auf k existiert ($\alpha \leq_k \beta$). In der Tabelle 5 sind die Definitionen der strukturellen Subsumption für einige termbildende Operatoren präsentiert.

3.1.2 Berechnung des LCS für zwei Konzepte

Es wurde gezeigt, wie die partielle Ordnungsrelation von einzelnen Konstrukten in den komplexen Beschreibungen definiert ist. Die ebenso strukturbasierten Definitionen des LCS-Operators für Konstruktionsoperatoren können in ähnlicher Manier aufgestellt werden. In der Tabelle 6 sind einige der LCS-Definitionen aufgezeichnet. Eine ausführlichere Liste findet man z.B. in [18]. Auch bei der Berechnung von LCS wird vorausgesetzt, daß die Ausgangskonzepte in Normalform vorliegen.

Um nur einige Eintragungen in der Tabelle zu kommentieren, betrachten wir die LCS-Methode für den „at-least" Operator. Wenn die Rollen gleich sind ($R_1 = R_2$), dann hängt die strukturelle Subsumptionsrelation nur von den ganzzähligen Werten der minimalen Anzahl der Rollenfüller (l_1 bzw. l_2) ab. Die Ordnungsrelation stellt in diesem konkreten Fall die $\geq$-Relation auf der Menge der ganzen Zahlen dar. Das Konzept $(\geq\ 3\ R)$ ist spezieller als $(\geq\ 1\ R)$, weil $3 \geq 1$ gilt. Die kleinste Anzahl der Rollenfüller, bei der die beiden Konzepte erfüllt bleiben, ist das Maximum von 3 und 1, also 3. Verallgemeinert folgt daraus, daß die Berechnung des LCS-Operators von $(\geq\ l_1\ R_1)$ und $(\geq\ l_2\ R_2)$ auf die Besimmung des Maximums von l_1 und l_2 hinausläuft, wenn die Rollen R_1 und R_2 gleich sind.

Operator k	LCS_k
and	$\text{LCS}((\text{and } C_1 \ldots, C_n), (\text{and } D_1 \ldots D_m)) = (\text{and LCS}(C_1, D_1) \ldots \text{LCS}(C_1, D_m) \ldots \text{LCS}(C_n, D_1) \ldots \text{LCS}(C_n, D_m))$
all	$\text{LCS}((\text{all } R_1\ C_1), (\text{all } R_2\ C_2)) = \begin{cases} (\text{all } R_1 \text{ LCS}(C_1, C_2)) & \text{wenn } R_1 = R_2 \\ \top & \text{sonst} \end{cases}$
at-least	$\text{LCS}((\text{at-least } l_1\ R_1), (\text{at-least } l_2\ R_2)) = \begin{cases} (\text{at-least } \mathit{min}(l_1, l_2)\ R_1) & \text{wenn } R_1 = R_2 \\ \top & \text{sonst} \end{cases}$
at-most	$\text{LCS}((\text{at-most } u_1\ R_1), (\text{at-most } u_2\ R_2)) = \begin{cases} (\text{at-most } \mathit{max}(u_1, u_2)\ R_1) & \text{wenn } R_1 = R_2 \\ \top & \text{sonst} \end{cases}$
atomic concept	$\text{LCS}(C_1, C_2) = \begin{cases} C_1 & \text{wenn } C_1 = C_2 \text{ und } C_i \sqsubseteq \top \\ \top & \text{sonst} \end{cases}$

Tabelle 6: Der LCS-Operator.

Beispiele In zwei weiteren Beispielen wird auf die TV-Sportsendungen-Wissensbasis zurückgegriffen (siehe den Abschnitt 2.3.5). Zur Erinnerung, die Wissensbasis beinhaltet folgende Konzeptdeklarationen:

$$FB \doteq TSB \sqcap (\geq\ 1\, HST) \sqcap (\leq\ 1\, HST) \sqcap \forall\, HST.F$$

$$BB \doteq TSB \sqcap (\geq\ 1\, HST) \sqcap (\leq\ 1\, HST) \sqcap \forall\, HST.B$$

$$TB \doteq ISB \sqcap (\geq\ 1\, HST) \sqcap (\leq\ 1\, HST) \sqcap \forall\, HST.TR$$

Im ersten Beispiel wird der LCS von FB und BB ermittelt. Als erstes werden die Konzepte normalisiert. Da es im vorliegenden Fall um zwei definierte Konzepte handelt, werden sie durch die Beschreibungen auf der rechten Seite der Konzeptdefinitionen ersetzt. Desweiteren werden die primitiven Konzepte TSB, F und BB laut Definition jeweils durch Konstrukte $(SB \sqcap TSB)$, $(ST \sqcap F)$ und $(ST \sqcap B)$ substituiert. Das ergibt den folgenden LCS-Operator:

$$\begin{aligned} &\text{LCS}(FB, BB) = \\ \text{LCS}\ \ &(\overbrace{SB \sqcap TSB \sqcap (\geq\ 1\, HST) \sqcap (\leq\ 1\, HST) \sqcap \forall\, HST.(ST \sqcap F)}^{FB}, \\ &\overbrace{SB \sqcap TSB \sqcap (\geq\ 1\, HST) \sqcap (\leq\ 1\, HST) \sqcap \forall\, HST.(ST \sqcap B)}^{BB}) \end{aligned}$$

Die Berechnung des LCSfür die einzelnen Teilterme führt zu dem Ergebnis:

$$\text{LCS}(FB, BB) = TSB \sqcap (\geq\ 1\, HST) \sqcap (\leq\ 1\, HST) \sqcap \forall\, HST.ST$$

Im Informationsrecherche-Kontext kann dieses Ergebnis durchaus als akzeptabel bewertet werden. Einerseits unterscheidet es sich von den Eingabekonzepten FB und BB, d.h. ist nicht zu speziell, andererseits ist es auch nicht zu allgemein.

Anders sieht das Ergebnis von LCS(FB, TB) aus:

$$\mathrm{LCS}(FB, TB) =$$

$$\mathrm{LCS}\ (\overbrace{SB \sqcap TSB \sqcap (\geq\ 1\, HST) \sqcap (\leq\ 1\, HST) \sqcap \forall\, HST.(ST \sqcap F)}^{FB},$$
$$\overbrace{SB \sqcap ISB \sqcap (\geq\ 1\, HST) \sqcap (\leq\ 1\, HST) \sqcap \forall\, HST.(ST \sqcap B)}^{TB}) =$$

$$SB \sqcap (\geq\ 1\, HST) \sqcap (\leq\ 1\, HST) \sqcap \forall\, HST.ST$$

Wie man leicht feststellen kann, beschreibt das resultierende LCS-Konzept in diesem zweiten Beispiel alle Sportsendungen mit einem Sportgerät. Das Konzept ist so allgemein, daß es sehr viele Individuen gibt, die Instanzen von dem Konzept sind. Das Ergebnis ingesamt muß deshalb als nicht geeignet für den Zweck der Informationsrecherche bewertet werden. Das Problem besteht darin, daß der speziellste gemeinsame Subsumierer von TSB und ISB das allgemeine Konzept SB ist. Das ist die Folge der vollständiger Subsumption, die hier vorausgesetzt wird. Wenn die Forderung einer vollständigen Subsumption allerdings aufgehoben wird, dann kann man hoffen von LCS bessere Ergebnisse zu erzielen. In dem betracheten Fall kann man sich unter „besseren Ergebnissen“ z.B. die folgenden Konzepte vorstellen:

$$ISB \sqcap (\geq\ 1\, HST) \sqcap (\leq\ 1\, HST) \sqcap \forall\, HST.ST$$
$$TSB \sqcap (\geq\ 1\, HST) \sqcap (\leq\ 1\, HST) \sqcap \forall\, HST.ST$$

3.1.3 Motivation für einen probabilistischen Ansatz

Die praktischen Ergebnisse der Anwendung der LCS-Methode im Kontext eines beispielorientierten Information Retrieval führten zur Erkenntnis, daß eine Verbesserung erwünscht ist. Eine der möglichen Lösungen ist der Verzicht auf die Forderung einer vollständigen Subsumption und das Heranziehen von Konzeptüberlappungen zur Berechnung von LCS-Konzeptkandidaten. Eine probabilistische Variante des LCS-Operators (genannt PLCS) wurde im Rahmen dieser Arbeit entwickelt.

Außerdem wurde mit Implementation von P-Classic eine Testumgebung für die Umsetzung und Verifikation des PLCS-Operators geschaffen. Die ausgearbeitete PLCS-Methode konnte in das implementierte PLCS-Rahmensystem integriert und am Beispiel einer konkreten Domain (TV-Sportsendungen) getestet werden.

In weiteren Abschnitten wird der PLCS-Operator ausführlich beschrieben und der Algorithmus zu Berechnung von PLCS-Konzeptkandidaten vorgestellt und an Beispielen erklärt.

3.2 Der probabilistische LCS-Operator

3.2.1 Definition des PLCS-Operators

Ähnlich zu dem beschriebenen LCS-Operator läßt sich ein „probabilistischer LCS“-Operator formlos als ein Operator beschreiben, der für eine Menge von Konzepten neue Konzeptbeschriebungen generiert, welche die Konzepte mit einer bestimmten positiven Wahrscheinlichkeit subsumieren. Um den Unterscheid zu dem LCS-Operator zu verdeutlichen, sollten zwei wichtige Stellen in dieser Definition hervorgehoben werden. Zum einem handelt es sich um eine probabilistische Subsumption, die auch nicht vollständige Konzeptüberlappung zuläßt.

Zum anderen kann unter Berücksichtigung der möglichen Konzeptüberlappungen durchaus mehr als ein PLCS-Konzept erzeugt werden. Im Gegensatz zum LCS ist das Ergebnis des PLCS-Operators nicht eindeutig. Deshalb stellt sich die nächste Frage nach der Bewertung von PLCS-Kandidaten. Hierfür berechnet der PLCS-Operator für jeden Konzeptkandidaten zwei Bewertungsmaße - „Recall“ und „Precision“ und wählt anschließend diejenigen Konzeptkandidaten aus, die den „optimalen“ „Recall“/„Precision“-Wert aufweisen. Auf die Bedeutung und Definition der Bewertungsmaße wird später eingegangen.

Um den PLCS-Operator formal einzuführen, muß zuerst gesagt werden, was ein „probabilistischer speziellster gemeinsamer Subsumierer“ ist. Die Definition wird in Begriffen der Wahrscheinlichkeitsrechnung gegeben.

$C_1, \ldots, C_m$ sind Konzepte in der kanonischen Form. P^* steht für die Wahrscheinlichkeitsverteilung aller Objekte der Wissensbasis (z.B. die Wurzel-P-Klasse einer P-Classic-Wissensbasis). Ein Konzept E wird als ein „probabilistic Least Common Subsumer“ der Konzepte $C_1, \ldots, C_m$ gekennzeichnet, wenn die folgenden probabilistischen terminologischen Axiome in Bezug auf P^* erfüllt sind:

$$P_{P^*}(E \wedge C_1) = \rho_1, \ldots, P_{P^*}(E \wedge C_m) = \rho_m,\ i \in [1, n]\ :\ \rho_1, \ldots, \rho_m > 0$$

Der PLCS-Operator von $C_1, \ldots, C_m$ berechnet eine Menge von Tripeln der folgenden Form: (E_i, r_{E_i}, p_{E_i}). Das erstes Element eines Tripels (E_i) ist ein „probabilistic Least Common Subsumer“-Konzept, r_{E_i} steht für Recall und p_{E_i} für Precision dieses Konzepts. Die nächste Bedingung ist, daß der PLCS-Operator anhand der Recall/Precision-Maße nur diejenigen Konzeptkandidaten übrigläßt, die über andere Kandidaten „dominieren“.

3.2.2 Dominanz

Seien E_1 und E_2 die PLCS-Kandidaten mit den Recall/Precision-Maßen r_{E_1}, p_{E_1} bzw. r_{E_2}, p_{E_2}. Das Tripel (E_1, r_{E_1}, p_{E_1}) *dominiert* über das Tripel (E_2, r_{E_2}, p_{E_2}), wenn $r_{E_1} \geq r_{E_2} \wedge p_{E_1} \geq p_{E_2}$ gilt.

Die formale Definition des PLCS-Operators sieht dann wie folgt aus:

$$\begin{aligned} &P\mathrm{LCS}(C_1, \ldots, C_m) = \\ &\quad \{(E_i, r_{E_i}, p_{E_i}),\ i \in [1, n] \mid \neg\exists E_j : E_j \text{ ist ein } p\text{-LCS und dominiert über} E_i, i \neq j\} \end{aligned}$$

3.2.3 Minimalität

Zum Schluß kann noch die Bedingung formuliert werden, unter der die Menge der PLCS-Kandidaten als minimal bezeichnet wird.

Sei $M = \{(E_1, r_{E_1}, p_{E_1}), \ldots, (E_n, r_{E_n}, p_{E_n})\}$ die Menge der PLCS-Kandidaten. Sie heißt *minimal*, wenn $\{\forall\ (E_i, r_{E_i}, p_{E_i}) \mid (E_j, r_{E_j}, p_{E_j}) \in M,\ i \neq j\ :\ E_i \not\equiv E_j\}$.

Im folgenden wird nun eine dreischrittige Methode zur Berechnung einer Menge von PLCS-Konzepten vorgestellt.

3.3 PLCS-Algorithmus

Die PLCS-Prozedur zur Berechnung von Konzeptkandidaten läßt sich aus grober Sicht in folgende drei Abläufe zerlegen:

$C_1, C_2, \ldots, C_m$ sind die Eingabekonzepte in der kanonischen Form.

1. Schritt: Finde alle PLCS-Kandidaten $E_1, \ldots, E_n$ für die Eingabekonzepte C_1, C_2, …, C_m;

2. Schritt: Für jeden ermittelten PLCS-Kandidaten E_i berechne die Bewertungsmaße Recall r_i und Precision $p_i : M = \{(E_i, r_i, p_i) \mid i \in [1, n]\}$

3. Schritt: Gegeben ist die Menge der Tripel $M = \{(E_i, r_i, p_i) \mid i \in [1, n]\}$. Finde dominierende PLCS-Kandidaten.

Unsere Betrachtungen widmen sich zuerst dem 1.Schritt des PLCS-Verfahrens.

3.3.1 Berechnung der PLCS-Konzeptkandidaten

Beschreibung des Algorithmus In diesem Abschnitt soll die Methode zur Berechnung von PLCS-Konzeptkandidaten erläutert werden. Abbildung 12 präsentiert den entwickelten und implementierten Algorithmus **Compute** PLCS**-Candidates** in einer universellen Programmiersprache.

Compute PLCS**-Candidates** $(\{C_1^d, \ldots, C_m^d\}, \mathit{pclass})$
Initialization: $M_1, \ldots, M_7 := \emptyset$, $\mathit{pclass} := P^*$, $d := \mathit{max}\{\mathit{depth}(C_1), \ldots, \mathit{depth}(C_m)\}$
  for $D \in \mathit{literals}$ (T-Box) **do**
    füge D in die Menge M_1 ein, wenn $\forall i \in \{1, \ldots, m\} : P_{\mathit{pclass}}(D \sqcap C_i) > 0$
  for $E := D_1 \sqcap \ldots \sqcap D_k \in 2^{M_1}$ **do**
    füge E zu der Menge M_2 hinzu, wenn $\forall i \in \{1, \ldots, m\} : P_{\mathit{pclass}}(E \sqcap C_i) > 0$
  for $R \in \mathit{roles}$ $(\{C_1, \ldots, C_m\})$ **do**
    for $(i, j) \in$ *number-restrictions* $(R, \{C_1, \ldots, C_m\})$ **do**
      füge $(\geq\ l\ R) \sqcap (\leq\ u\ R)$ zu der Menge M_3 hinzu,
      wenn $0 \leq l \leq j \wedge u \leq i$ $\mathit{bound}(R) \wedge l \leq u$
    for $\{C_1^{d-1}, \ldots, C_m^{d-1}\} \in$ *all-quantification* $(R, \{C_1, \ldots, C_m\})$ **do**
      for $\mathit{rpclass} \in$ *relevant-p-classes* $(\{C_1^{d-1}, \ldots, C_m^{d-1}\}, \{C_1, \ldots, C_m\})$ **do**
        füge Ergebnisse von
        Compute PLCS**-Candidates**$(\{C_1^{d-1}, \ldots, C_m^{d-1}\}, \mathit{rpclass})$
        zu der Menge M_4 hinzu
    for $E := A \sqcap \forall R.B : A \in M_3, B \in M_4$ **do**
      nimm E in die Menge M_5 auf
  for $E := D_1 \sqcap \ldots \sqcap D_k \in 2^{M_5}$ **do**
    nimm E in die Menge M_6 auf, wenn $\forall i \in \{1, \ldots, m\} : P_{\mathit{pclass}}(E \sqcap C_i) > 0$
  for $E := A \sqcap B : A \in M_2, B \in M_6$ **do**
    füge E zu der Menge M_7 hinzu, wenn $\forall i \in \{1, \ldots, m\} : P_{\mathit{pclass}}(E \sqcap C_i) > 0$
  gib $M_2 \cup M_7$ zurück

Abbildung 12: Der PLCS-Algorithmus **Compute** PLCS**-Candidates**.

Jetzt wird auf die einzelnen Zwischenschritte der PLCS-Methode ausführlicher eingegangen.

$\{C_1, \ldots, C_m\}$ sind Konzepte der maximalen Tiefe d. Damit der Algorithmus terminiert, wird vorausgesetzt, daß die maximale Tiefe der PLCS-Konzepte ebenfalls einen endlichen Wert d hat. In praktischen Anwendungen bedeutet dies keine bedeutende Einschränkung, denn je mehr ein Konzept verschachtelt ist, desto unwahrscheinlicher ist, daß es Instanzen hat (z.B. $TSB \sqcap \forall\, HST.\forall\, HST.F$).

1. Am Anfang ist die Variable *pclass* mit der Wurzel-P-Klasse einer *P*-CLASSIC-Wissensbasis belegt. Die Menge M_1 besteht aus atomaren Konzepten der T-Box, die sich in der P-Klasse *pclass* mit $C_1, \ldots, C_m$ überlappen. Im folgenden wird unter Überlappung die probabilistische Subsumption in Bezug auf eine gegebene P-Klasse *pclass* verstanden. Die Funktion *literals* liefert die Menge aller atomaren bzw. negierten atomaren Konzepte der Terminologie.

2. Die Menge M_2 enthält von den maximal 2^{M_1} möglichen Konjunktionen der Konzepte aus der Menge M_1 diejenigen, die sich mit $C_1, \ldots, C_m$ überlappen.

3. Die Menge M_3 enthält für jede Rolle der Wissensbasis die Anzahlrestriktionen der Rollenfüller. Für welche Rollen neue Constraints gebaut werden müssen, bestimmt die Funktion *roles*. Mit Hilfe der Funktion *number-restrictions* wird für jede Rolle die „Spannweite" der Fülleranzahl berechnet. Liegt beispielsweise die Zahl der Füller der Rolle R in C_1 zwischen 1 und 3 und in C_2 zwischen 2 und 5, so gibt die Prozedur *number-restrictions* (R, $\{C_1, C_2\}$) das Intervall [1,5] zurück. Die erzeugten Constraints müssen sich mit den in $C_1, \ldots, C_m$ vorliegenden Intervallen überschneiden. Kommt in einem Konzept C_i eine Rolle R_j gar nicht vor, so gilt bzgl. der Anzahl der Rollenfüller der Rolle R_j die Einschränkung [0, *bound*(R_j)]. Da die maximale Anzahl der Rollenfüller für jede Rolle als eingeschränkt angenommen wird, ist die Menge M_3 endlich.

4. Die Menge M_4 entsteht nach dem rekursiven Aufruf des Algorithmus mit Konzepttermen der Tiefe d-1. Diese Konzepte beschreiben Eigenschaften der Rollenfüller. Die Aufgabe der Funktion *relevant-p-classes* besteht darin, für Konzeptterme der Tiefe d-1 und somit für Rollenfüller die richtigen Wahrscheinlichkeitsverteilungen (P-Klassen) zu bestimmen. Eine P-Klasse wird als relevant für ein Konzept C bezeichnet, wenn die Wahrscheinlichkeit des Konzepts C in dieser P-Klasse nicht 0 ist.

5. Die Elemente der Menge M_5 sind aus den Anzahlrestriktionen der Menge M_3 und aus den qualifizierten Werteinschränkungen aus M_4 zusammengebaut. Dabei müssen alle Rollen der Wissensbasis berücksichtigt werden.

6. Die Menge M_6 besteht aus allen möglichen Konjunktionen der Konzeptterme aus der Menge M_5 (höchstens 2^{M_5} Terme), die sich mit $C_1, \ldots, C_m$ überlappen.

7. Die Menge M_7 enthält von allen möglichen Konjunktionen der M_2- und M_6-Konzeptterme diejenigen Konzepte, die sich mit $C_1, \ldots, C_m$ überlappen.

8. Das Ergebnis ist die Vereinigung der Mengen M_2, M_6 und M_7.

Beispiel Zum Vergleich der Ergebnisse der vorgestellten *P*LCS-Prozedur mit dem Resultat der nicht-probabilistischen LCS-Methode wird das im Abschnitt 3.1.2 diskutierte Beispiel erneut aufgregriffen. Die Aufgabe besteht darin, die *P*LCS-Funktion der Konzepte `football-broadcast` (FB) und `tennis-broadcast` (TB) zu berechnen:

$$FB = TSB \sqcap (\geq\ 1\, HST) \sqcap (\leq\ 1\, HST) \sqcap \forall\, HST.F$$

$$TB = ISB \sqcap (\geq\ 1\, HST) \sqcap (\leq\ 1\, HST) \sqcap \forall\, HST.TR$$

Die beiden Konzepte haben die Tiefe $d = 1$. Die beschriebene Prozedur **Compute** $\mathcal{P}$LCS**-Candidates** fängt mit den Teilausdrücken der maximalen Tiefe d an und wird rekursiv für die Teilasdrücke der Tiefe d - 1 aufgerufen. Der zweite Aufrufparameter ist eine P-Klasse und wird am Anfang mit der Wurzel-P-Klasse „sports-broadcasts" initiiert. In den späteren rekursiven Aufrufen wird sie nach der erwähnten Methode *relevant-p-classes* bestimmt. Die Prozedur **Compute** $\mathcal{P}$LCS**-Candidates** gibt in unserem Beispiel die folgenden Zwischenergebnisse zurück.

- Als erstes werden atomare Konzepte der T-Box gesucht, die sich mit den Konzepten FB und TB in der Wurzel-P-Klasse „sports-broadcasts" überlappen. Zur Vereinfachung werden negierte atomare Konzepte in diesem Beispiel nicht berücksichtigt. Die Terminologie der Beispielwissensbasis für Sportsendungen enthält folgende atomare Konzepte: $\{SB, TSB, ISB, ST, B, F, TR\}$ (vgl. 2.3.5). Die Berechnung der probabilistischen Subsumptionsrelation zwischen FB bzw. TB und den jeweiligen atomaren Konzepten liefert im Ergebnis die Menge M_1:

$$M_1 = \{SB, TSB, ISB\}$$

- Danach werden aus Elementen der Menge M_1 alle möglichen Konjunktionen gebildet (inklusive die Ein-Konjunkt-Terme) und die neuen Konzepte wieder mit FB bzw. TB hinsichtlich der Überlappung verglichen. Die resultierende Menge M_2 beinhaltet die folgenden Konstrukte:

$$M_2 = \{SB, TSB, ISB, SB \sqcap TSB, SB \sqcap ISB, TSB \sqcap ISB, SB \sqcap TSB \sqcap ISB\}$$

- Im nächsten Berechnungschritt werden Constraints für die Anzahl der Rollenfüller erzeugt, die sich mit den in FB bzw. TB spezifizierten Constraints überlappen. Es handelt sich um die Rolle HST, deren Grenzwerte für die Anzahl der Füller in der Wissensbasis bei 0 bzw. 1 liegen. Die möglichen Wertintervalle sind [0,0], [0,1] und [1,1]. In den Konzeptdefinitionen von FB und TB ist die Anzahl der Rollenfüller noch zusätzlich eingeschränkt: $(\geq\ 1\, HST) \sqcap (\leq\ 1\, HST)$. Man kann leicht verifizieren, daß nur [0,1] und [1,1] sich mit den vorgegebenen Constraints überschneiden. Damit ergibt sich die folgende Menge $M_3 = \{(\geq\ 0\, HST) \sqcap (\leq\ 1\, HST), (\geq\ 1\, HST) \sqcap (\leq\ 1\, HST)\}$. Der Übersichtlichkeit wegen werden im folgenden für die ermittelten Constraints die Kurzbezeichnungen HST_0^1 bzw. HST_1^1 verwendet:

$$M_3 = \{HST_0^1, HST_1^1\}$$

- Als nächstes folgt der rekursive Aufruf der $\mathcal{P}$LCS-Prozedur für die $\forall$-Quantifi-katoren der Rolle HST. Das sind die Konzeptterme F und TR der Tiefe 0. Erneut werden in der T-Box atomare Konzepte gesucht, die sich mit F und TR in Bezug auf die entsprechende(n) Wahrscheinlichkeitsverteilung(en) (P-Klasse(n)) überlappen. Hierfür müssen die relevanten P-Klassen, wie im Abschnitt 3.3.1 erklärt, bestimmt werden. In unserem Beispiel sind das die beiden P-Klassen „sports-broadcasts" und „sports-tools". Nur das Konzept ST subsumiert F und TR in beiden P-Klassen und wird deshalb in die Menge M_4 aufgenommen:

$$M_4 = \{ST\}$$

- Die Menge M_5 setzt sich aus allen möglichen Konjunktionen von ermittelten Zahlen- und Wertrestriktionen (gesammelt in den Mengen M_3 und M_4) für die Füller der Rolle HST zusammen:

$$M_5 = \{HST_0^1 \sqcap \forall\, HST.ST, HST_1^1 \sqcap \forall\, HST.ST\}$$

- Aus Elementen der Menge M_5, werden alle möglichen Konjunktionen zusammengebaut, in denen keine Rolle mehr als einmal vorkommt und die mit den Konzepten FB und TB in der Wurzel-P-Klasse überlappen. In unserem Beispiel gibt es nur eine Rolle, deshalb ist die Menge M_6 identisch mit M_5:

$$M_6 = M_5$$

- Der letzte Schritt besteht in der Konstruktion von allen möglichen Konjunktionen der Konzeptterme aus der Menge M_2 (das sind atomare Konzepte und Konjunktionen von atomaren Konzepten) und der Menge M_6 (Anzahl- und Werteinschränkungen der Rollenfüller). Auch in diesem Fall gilt die Bedingung, daß die neuen Konzepte sich mit den Eingabekonzepten FB und TB in der Wurzel-P-Klasse „sports-broadcasts“ überlappen. Die Ergebnisse dieses Vorgangs werden als Menge M_7 zusammengefasst. Letztendlich stellt die Vereinigung von M_7 und M_2 das Resultat des Ausführens von **Compute** *P*LCS**-Candidates** dar. Die resultierende Menge besteht aus 21 neuen Konzepttermen, den sogenannten *P*LCS-Kandidaten:

$$\begin{aligned}
&M_2 \cup M_7 = \\
&\quad \{SB, TSB, ISB, SB \sqcap TSB, SB \sqcap ISB, TSB \sqcap ISB, SB \sqcap TSB \sqcap ISB, \\
&\quad SB \sqcap HST_0^1 \sqcap \forall\, HST.ST, \\
&\quad SB \sqcap HST_1^1 \sqcap \forall\, HST.ST, \\
&\quad TSB \sqcap HST_0^1 \sqcap \forall\, HST.ST, \\
&\quad TSB \sqcap HST_1^1 \sqcap \forall\, HST.ST, \\
&\quad ISB \sqcap HST_0^1 \sqcap \forall\, HST.ST, \\
&\quad ISB \sqcap HST_1^1 \sqcap \forall\, HST.ST, \\
&\quad SB \sqcap TSB \sqcap HST_0^1 \sqcap \forall\, HST.ST, \\
&\quad SB \sqcap TSB \sqcap HST_1^1 \sqcap \forall\, HST.ST, \\
&\quad SB \sqcap ISB \sqcap HST_0^1 \sqcap \forall\, HST.ST, \\
&\quad SB \sqcap ISB \sqcap HST_1^1 \sqcap \forall\, HST.ST, \\
&\quad TSB \sqcap ISB \sqcap HST_0^1 \sqcap \forall\, HST.ST, \\
&\quad TSB \sqcap ISB \sqcap HST_1^1 \sqcap \forall\, HST.ST, \\
&\quad SB \sqcap TSB \sqcap ISB \sqcap HST_0^1 \sqcap \forall\, HST.ST, \\
&\quad SB \sqcap TSB \sqcap ISB \sqcap HST_1^1 \sqcap \forall\, HST.ST\}
\end{aligned}$$

Die Analyse der erhaltenen Ergebnisse zeigt, daß die *P*LCS-Konzepte die Eingabekonzepte auf ganz unterschiedliche Art und Weise überlappen. Das Konzept SB subsumiert sie zum Beispiel vollständig und ist zu allgemein. Es ist notwendig, einen Weg zu finden, die „Qualität“ der ermittelten Konzeptkandidaten messen zu können und nur die mit den „optimalen“ Parametern auszuwählen. Für die Lösung dieser Aufgabe werden zwei Bewertungsmaße herangezogen. Der Ansatz orientiert sich an die Performanzmessung in den traditionellen Information Retrieval Systemen. Die Möglichkeit dazu bieten u.a. Recall/Precision-basierte Maße. Die Parameter, die für die Bewertung der PLCS-Kandidaten verwendet werden, sind den Recall/Precision-Maßen ziemlich ähnlich, beziehen sich aber auf die probabilistische Subsumption.

3.3.2 Methode zur Bewertung der PLCS-Kandidaten

Im Kontext der traditionellen Informationsrecherche sind folgende Definitionen von Recall und Precision üblich (siehe z.B. [21] oder [19]) Recall ist die Wahrscheinlichkeit, daß ein relevantes Dokument vom System als solches erkannt wird. Precision ist die Wahrscheinlichkeit, daß ein vom System als relevant erkanntes Dokument in der Tat relevant ist. Die beiden Maße liegen im Wertebereich [0,1].

Die Definitionen für Recall und Precision werden im P-Classic-Kontext wie folgt formuliert.

Recall $C_1, \ldots, C_m$ sind Eingabekonzepte. E ist ein PLCS-Konzeptkandidat. Das Recall-Maß läßt sich nach der folgenden Formel berechnen:

$$r_{E,C_1,\ldots,C_m} = P(C_1 \cup \ldots \cup C_2 | E) = \frac{P(E \cap (C_1 \cup \ldots \cup C_m))}{P(E)}$$

Recall spiegelt die bedingte Wahrscheinlichkeit wider, mit der ein zufällig gewähltes Objekt der Anwendungsdomäne eine Instanz der Konzepte $C_1, \ldots, C_m$ ist, wenn bekannt ist, daß es eine Instanz des PLCS-Konzepts E ist.

Precision $C_1, \ldots, C_m$ sind Eingabekonzepte. E ist ein PLCS-Konzeptkandidat. Das Precision-Maß drückt die Wahrscheinlichkeit aus, mit der ein zufällig gewähltes Objekt der Anwendungsdomäne eine Instanz von E ist, wenn bekannt ist, das es auch eine Instanz von jedem der $C_1, \ldots, C_m$ Konzepte ist. Die Definition sieht entsprechend folgendermassen aus:

$$p_{E,C_1,\ldots,C_m} = P(E | C_1 \cup \ldots \cup C_2) = \frac{P(E \cap (C_1 \cup \ldots \cup C_m))}{P(C_1 \cup \ldots \cup C_m)}$$

In beiden Berechnungsvorschriften wird auf die wahrscheinlichkeitstheoretische Formel zur Berechnung der Wahrscheinlichkeit einer Summe von zufälligen Ereignissen zurückgegriffen (s. Abschnitt 2.4.2).

Beispiel Nachdem deutlich geworden ist, wie die Bewertungsmaße ermittelt werden, wollen wir die Berechnung von PLCS-Kandidaten aus dem Beispiel im letzten Abschnitt fortsetzen. Für die ermittelten Konzeptkandidaten werden nun die Precision/Recall-Werte ausgerechnet. Das Ergebnis stellt die folgende Menge von 21 Tripeln dar:

$$\begin{aligned}
M = & \\
& \{(SB, 0.2, 1.0), \\
& (TSB, 0.22, 0.35), \\
& (ISB, 0.23, 0.35),
\end{aligned}$$

$(SB \sqcap TSB, 0.22, 0.35),$
$(SB \sqcap ISB, 0.23, 0.35),$
$(TSB \sqcap ISB, 0.23, 0.14),$
$(SB \sqcap TSB \sqcap ISB, 0.23, 0.14),$
$(SB \sqcap HST_0^1 \sqcap \forall HST.ST, 0.22, 1.0),$
$(SB \sqcap HST_1^1 \sqcap \forall HST.ST, 0.22, 1.0),$
$(TSB \sqcap HST_0^1 \sqcap \forall HST.ST, 0.24, 0.35),$
$(TSB \sqcap HST_1^1 \sqcap \forall HST.ST, 0.24, 0.35),$
$(ISB \sqcap HST_0^1 \sqcap \forall HST.ST, 0.26, 0.35),$
$(ISB \sqcap HST_1^1 \sqcap \forall HST.ST, 0.26, 0.35),$
$(SB \sqcap TSB \sqcap HST_0^1 \sqcap \forall HST.ST, 0.24, 0.35),$
$(SB \sqcap TSB \sqcap HST_1^1 \sqcap \forall HST.ST, 0.24, 0.35),$
$(SB \sqcap ISB \sqcap HST_0^1 \sqcap \forall HST.ST, 0.26, 0.35),$
$(SB \sqcap ISB \sqcap HST_1^1 \sqcap \forall HST.ST, 0.26, 0.35),$
$(TSB \sqcap ISB \sqcap HST_0^1 \sqcap \forall HST.ST, 0.26, 0.14),$
$(TSB \sqcap ISB \sqcap HST_1^1 \sqcap \forall HST.ST, 0.26, 0.14),$
$(SB \sqcap TSB \sqcap ISB \sqcap HST_0^1 \sqcap \forall HST.ST, 0.26, 0.14),$
$(SB \sqcap TSB \sqcap ISB \sqcap HST_1^1 \sqcap \forall HST.ST, 0.26, 0.14)\}$

Bedeutung von Recall und Precision in PLCS-Kontext Das Recall-Maß soll den Grad der Subsumption von Eingabekonzepten $C_1, \ldots, C_m$ durch den gegebenen PLCS-Kandidaten E zum Ausdruck bringen. Je größer der Recall-Wert ist, desto weniger Instanzen des Konzepts E nicht auch die Instanzen von $C_1, \ldots, C_m$ sind. Einen perfekten Recall-Wert hat z.B. ein Konzept E, das genau die Eingabekonzepte subsumiert. Der Nachteil ist, daß E keine neuen Individuen beschreibt, die nicht schon von Eingabekonzepten beschrieben wären. Im Informationsrecherche-Kontext folgt daraus, daß keine Alternativen ermittelt werden, sollte das Konzept E im Retrieval-Inferenzdienst verwendet werden. Das Konzept SB hat z.B. das schlechteste Recall-Maß, weil es zu allgemein ist. Anhand der Recall-Werte können deshalb die zu allgemeinen PLCS-Konzeptkandidaten ausgeschlossen werden.

Andererseits ist die vollständige Subsumption auch keine Bedingung für die PLCS-Berechnung. Deshalb wird ein zweiter Faktor, Precision, herangezogen, der zeigen soll, wie genau die Überlappung wirklich ist. Wenn sich herausstellt, daß die meisten Instanzen von $C_1, \ldots, C_m$ keine Instanzen des PLCS-Konzepts E sind, dann hat E einen schlechten Precision-Wert. Im besten Fall werden die Eingabekonzepte von E vollständig subsumiert. So hat der PLCS-Konzeptkandidat SB den höchsten Precision-Wert. Aber wir haben gesehen, daß er eigenltich viel zu allgemein ist. Andere Konzeptkandidaten haben keinen so guten Precision-Wert. Dafür ist der Recall-Wert bei ihnen größer und die Anzahl der Individuen, die Instanzen dieser Konzepte sind, entsprechend kleiner.

Um den Sinn der beiden Bewertungsmaße zum Abschluß dieser Diskussion noch einmal graphisch zu verdeutlichen, zeigt Abbildung 13 zwei Eingabekonzepte C_1, C_2 und zwei PLCS-Konzeptkandidaten E_1, E_2, so daß die Größe der Konzeptüberlappung mit Hilfe von unterschiedlich gemusterten Flächen visualisiert ist.

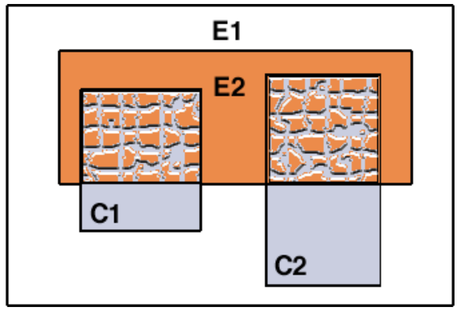

Abbildung 13: Illustration der Bedeutung der Recall/Precision-Maße.

Das Verhältnis der zweifarbigen Fläche zu der Gesamtfläche von E_2 spiegelt den Recall-Wert des E_2-Kandidaten wider. Je kleiner der Anteil der Fläche eines PLCS-Kandidaten ist, die sich weder mit C_1 noch mit C_2 überlappt, desto größer ist der Recall-Wert des Kandidaten. Umgekehrt, je weniger Fläche von C_1 und C_2 außerhalb der Fläche eines PLCS-Kandidaten liegt, desto besser ist die Precision-Charakteristik dieses Kandidaten. Zum Beispiel hat E_1 einen besseren Precision-Wert als E_2, weil C_1 und C_2 in E_1 vollständig enthalten sind. Dafür hat E_2 einen größeren Recall-Wert, weil der Anteil seiner Fläche, der außerhalb von C_1 und C_2 liegt, kleiner ist, als die weiße Fläche des Kandidaten E_1.

Der nächste Schritt der PLCS-Methode führt zur Bestimmung von optimalen PLCS-Kandidaten.

3.3.3 Bestimmung der optimalen PLCS-Kandidaten

Nach der Erzeugung und Bewertung von PLCS-Kandidaten folgt die Auswahl von „besten“ PLCS-Kandidaten. Die Grundlage hierfür bieten die ermittelten Recall und Precision Zahlen und die im Abschnitt 3.2.1 definierte Dominanz-Relation. Der Algorithmus zur Bestimmung von optimalen PLCS-Kandidaten zeigt Abbildung 14.

Compute Dominated PLCS**-Candidates** $((E_1, r_1, p_1), \ldots, (E_n, r_n, p_n))$
Initialization: $p\text{-}lcs(C_1, \ldots, C_m) = \{(E_i, r_i, p_i), i \in [1, n]\}$
  for $i = 1$ bis n **do**
    entferne alle Tripel (E_j, r_j, p_j) aus $p\text{-}lcs(C_1, \ldots, C_m)$, für die gilt:
    $r_j \leq r_i \wedge p_j \leq p_i$

Abbildung 14: Der Algorithmus **Compute Dominated** PLCS**-Candidates**.

Beispiel Die Berechnung von PLCS-Kandidaten für die Ausgangskonzepte FB und TB wird mit der Auswahl der dominierenden Tripel E_i, r_i, p_i nach dem Verfahren **Compute Dominated** PLCS**-Candidates** abgeschlossen. Das endgültige Ergebnis stellt eine Menge von 14 Konzeptkandidaten dar:

$$
\begin{aligned}
&P\text{LCS}(FB, TB) = \\
&\quad \{(SB \sqcap HST_0^1 \sqcap \forall\, HST.ST, 0.22, 1.0), \\
&\quad (SB \sqcap HST_1^1 \sqcap \forall\, HST.ST, 0.22, 1.0), \\
&\quad (TSB \sqcap HST_0^1 \sqcap \forall\, HST.ST, 0.24, 0.35), \\
&\quad (TSB \sqcap HST_1^1 \sqcap \forall\, HST.ST, 0.24, 0.35), \\
&\quad (ISB \sqcap HST_0^1 \sqcap \forall\, HST.ST, 0.26, 0.35), \\
&\quad (ISB \sqcap HST_1^1 \sqcap \forall\, HST.ST, 0.26, 0.35), \\
&\quad (SB \sqcap TSB \sqcap HST_0^1 \sqcap \forall\, HST.ST, 0.24, 0.35), \\
&\quad (SB \sqcap TSB \sqcap HST_1^1 \sqcap \forall\, HST.ST, 0.24, 0.35), \\
&\quad (SB \sqcap ISB \sqcap HST_0^1 \sqcap \forall\, HST.ST, 0.26, 0.35), \\
&\quad (SB \sqcap ISB \sqcap HST_1^1 \sqcap \forall\, HST.ST, 0.26, 0.35), \\
&\quad (TSB \sqcap ISB \sqcap HST_0^1 \sqcap \forall\, HST.ST, 0.26, 0.14), \\
&\quad (TSB \sqcap ISB \sqcap HST_1^1 \sqcap \forall\, HST.ST, 0.26, 0.14), \\
&\quad (SB \sqcap TSB \sqcap ISB \sqcap HST_0^1 \sqcap \forall\, HST.ST, 0.26, 0.14), \\
&\quad (SB \sqcap TSB \sqcap ISB \sqcap HST_1^1 \sqcap \forall\, HST.ST, 0.26, 0.14)\}
\end{aligned}
$$

Aufgrund der ermittelten Ergebnisse können folgende Beobachtungen gemacht werden.

Konzeptterme, die sich nur in der Anzahlrestriktionen der Rollenfüller unterscheiden (z.B. $TSB \sqcap HST_0^1 \sqcap \forall HST.ST$ bzw. $TSB \sqcap HST_1^1 \sqcap \forall HST.ST$) haben die gleichen Maßwerte. Die Erklärung dafür liegt in der konkreten Spezifikation unserer Wissensbasis. Es wurde festgelgt, daß ein Sportsendung-Objekt mit genau einem Füller in der Beziehung `has-sports-tool` steht (s. Abschnitt 2.4.5). Alle anderen Werte der Anzahl der Rollenfüller (auch 0) gelten für die Rolle `has-sports-tool` als unwahrscheinlich und haben keine Auswirkung auf den Wahrscheinlichkeitswert einer Aussage.

Das Konzept $SB \sqcap HST_1^1 \sqcap \forall\, HST.ST$ hat erwartungsgemäß den höchsten Precision-Wert 1. Da dieses Konzept auch ein LCS-Kandidat ist, werden die Eingabekonzepte von ihm vollständig subsumiert. Im Vergleich zu diesem Konzept haben die Kandidaten $TSB \sqcap HST_1^1 \sqcap \forall HST.ST$ und $ISB \sqcap HST_1^1 \sqcap \forall HST.ST$ eine schlechtere Precision-Charakteristik, aber ein größeres Recall-Maß. Sie überlappen die Eingabekonzepte zwar nicht vollständig, dafür ist die Anzahl der Anwendungsobjekte, die Instanzen dieser Konzepte sind, kleiner als bei dem LCS-Konzept. Für die Zwecke der Informationsrecherche stellen sie deshalb im Vergleich zu dem LCS-Konzept bessere Kandidaten dar.

3.4 PLCS-Probleme und Diskussion

Im Gegensatz zu der LCS-Methode kann der *P*LCS-Algorithmus nicht mehr in polynomialer Zeit ausgeführt werden. Die Komplexität der Berechnung wächst exponentiell zu der Größe der Wissensbasis. Analysiert man die Zwischenschritte des *P*LCS-Verfahrens (vgl. Abschnitt 3.3.1), so stellt man fest, daß im schlimmsten Fall 2^m Vergleiche von Eingabekonzepten mit den *P*LCS-Kandidaten hinsichtlich der probabilistischen Subsumption erforderlich sind. Der Parameter m steht hier für die Anzahl von atomaren und negierten atomaren Konzepten einer *P*-Classic-Wissensbasis. Jeder Vergleich führt zur Berechnung einer probabilistischen Inferenz in Bayes-Netzen der P-Klassen. Je mehr Konzeptüberlappungen in der Wissensbasis modelliert sind, desto aufwendiger gestaltet sich die Ermittlung von *P*LCS-Konzeptkandidaten. Die Optimierung der beschriebenen *P*LCS-Prozedur ist eine Voraussetzung für den Einsatz des Verfahrens in großen Wissensbasen mit vielen sich überlappenden Konzepten.

Bei der Informationsrecherche ist es oft das Ziel, nicht nur interessante Informationen zu finden, sondern auch den Anteil nicht interessanter Daten zu minimieren. Wenn die Recherche beispielorientiert ist, dann könnte das Szenario wie folgt aussehen. Der Benutzer wählt für seine Sammlung neben positiven auch die sogenannten negativen Beispiele aus. Das Informationsrecherche-System muß dann sicherstellen, daß ermittelte Objekte, die dem Benutzer präsentiert werden, möglichst wenig Gemeinsames mit den negativen Beispielen haben. Es liegt nahe, auch für diese Zwecke den *P*LCS-Operator zu verwenden. Allerdings muß seine Definition leicht angepaßt werden. Z.B. kann ein neues Maß zu Bewertung des *P*LCS-Operators in Bezug auf negative Beispiele eingeführt werden. Das ist auf jeden Fall ein interessantes Ziel für die weiteren Untersuchungen in diesem Bereich.

4 Zusammenfassung

Motiviert durch einen intelligenten benutzeradaptiven Informationsdienst wurden im Rahmen der vorliegenden Arbeit Techniken erforscht, mit denen die Leistungsfähigkeit einer Informationsrecherche gesteigert werden kann.

Als Ausgangspunkt für die theoretischen und praktischen Weiterentwicklungen in dieser Arbeit diente die Idee der beispielorientierten Informationsrecherche. In [13] wurden die wesentlichen Präsentations- und Abstraktionsmittel eingeführt, mit denen die Umsetzung dieser Idee in die Praxis realisiert werden kann. Dazu gehören Beschreibungslogiken und Verallgemeinerungsoperatoren. Beschreibungslogiken stellen einen Basisformalismus zur Modellierung von Objekten einer Domäne in Form von Konzepthierarchien dar. Mit Hilfe von Verallgemeinerungsoperatoren werden die Gemeinsamkeiten von Objekten (Beispielen) berechnet.

Ein Problem mit Beschreibungslogiken, das in der Forschung erkannt wurde, besteht darin, daß sie unabhängig von ihrer Mächtigkeit die Subsumptionsbeziehung nur qualitativ beschreiben. So kann die Disjunktheit, Äquivalenz oder grundsätzliche Subsumption von Konzepten ausgedrückt werden. Wenn eine Schlußfolgerung besagt, daß ein Konzept von einem anderen Konzept subsumiert wird, dann bedeutet dies zwangsläufig eine „vollständige“ Subsumption. Das Ergebnis eines Verallgemeinerungsoperators ist unter diesen Umständen oft entweder zu allgemein für den Zweck der Informationssuche oder findet keine neuen Objekte außer den bereits ausgewählten Beispielen.

Wenn allerdings ein Repräsentationsmittel gefunden wird, mit dem die Konzeptsubsumption auch quantitativ beschrieben werden kann, dann bestehen auch gute Aussichten auf ein akzeptables Ergebnis des Verallgemeinerungsoperators.

In dieser Arbeit wurde gezeigt, wie die Kombination einer Beschreibungslogik mit dem Bayes-Netz-Formalismus zur Lösung des beschriebenen Problems beitragen kann. Als Basis für die Untersuchungen wurde *P*-Classic, herangezogen. *P*-Classic wurde in [7] als eine Ergänzung der Beschreibungslogik Classic um probabilistische Inferenzdienste entwickelt. Die Grundfunktionalität von *P*-Classic wurde implementiert und verifiziert. Das Ergebnis stellt ein abgeschlossenes Rahmensystem dar, mit dem weitere Experimente und Forschungsarbeiten möglich sind.

Die Eigenschaft von *P*-Classic, den Grad der Überlappung von Konzepten statistisch auszudrücken, wurde für die Ausarbeitung eines probabilistischen Verallgemeinerungsoperators (*P*LCS) ausgenutzt. Der Algorithmus zur Berechnung des *P*LCS-Operators wurde vorgestellt und in das *P*-Classic-Wissenspräsentationssystem als zusätzlicher Inferenzdienst eingeführt. Es wurde an Beispielen gezeigt, daß mit dem Einsatz des *P*LCS-Operators im Kontext der Informationsrecherche bessere Ergebnisse erzielt werden als sie vom LCS-Operator zu erwarten sind. Im Gegensatz zu dem LCS-Operator, der die Überlappungsgröße eines erzeugten LCS-Konzepts in Bezug auf die Eingabekonzepte nicht berücksichtigen kann, wird von dem *P*LCS-Operator hierfür diverse quantitative Maße berechnet.

Die in den Abschnitten 2.5 und 3.4 diskutierten Probleme und Verbesserungsmöglichkeiten in Bezug auf den ersten Prototyp des *P*-Classic-Wissenspräsentationssystems und auf den *P*LCS-Algorithmus geben Ansatzpunkte für die künftigen Entwicklungen. Ohne auf die einzelnen Details noch einmal einzugehen, lassen sich zum Abschluß nur die wichtigsten Feststellungen wiederholen. In *P*-Classic spielt die Annahme der Unabhängigkeit von Rollenfüllern eine wichtige Voraussetzung für die Traktabilität der probabilistischen Inferenzen. Sie bedeutet aber auch entscheidende Einschränkung für die Ausdrucksmächtigkeit dieser Logik. Es sollte nach Möglichkeit gesucht werden, diese Einschränkung aufzuheben. Zu den Zielen der Weiterentwicklung von P-CLASSIC können außerdem die Optimierung des Inferenzverfahrens und die Verwendung einer ausdruckmächtigeren terminologischen Komponente gehören.

Obwohl die Ergebnisse des PLCS-Operators als zufriedenstellend bewertet werden können, ist die Berechnung des PLCS-Operators noch sehr aufwendig und bedarf der Entwicklung weiterer Optimierungstechniken. Im Kontext des Information Retrieval werden oft nicht nur für den Benutzer interessante Informationen gesucht, sondern müssen u.U. bestimmte Informationen vermieden werden. Der PLCS-Operator läßt sich leicht für diese Art von negativen Beispielen in Informationsanfragen ausbauen

Literatur

[1] A. Borgida. Description logics are not just for the flightless-birds: A new look at the utility and foundations of description logics. Research report, Department if Computer Science, Rutgers University New Brunswick, 1992.

[2] R. Brachman, D. McGuinness, P. F.Patel-Schneider, L. Alperin Resnick, and A. Borgida. Living with CLASSIC: When and how to use a KL-ONE-like language. In John F. Sowa, editor, *Principles of Semantic Networks — Explorations in the Representation of Knowledge*, pages 401–456. Morgan Kaufmann, 1991.

[3] W. Cohen and H. Hirsh. Learning the CLASSIC description logic: Theoretical and experimental results. In Pietro Torasso Jon Doyle, Erik Sandewall, editor, *Proc. of the 4th Int. Conf. on Principles of Knowledge Representation and Reasoning*. Morgan Kaufmann, May 1994.

[4] W. W. Cohen, A. Borgida, and H. Hirsh. Computing Least Common Subsumers in Description Logics. In *Proceedings of the National Conference on Artificial Intelligence AAAI'92*, pages 754–760. AAAI Press/The MIT Press, 1992.

[5] Yen J., Neches R., and Mac Gregor R. Using terminological models to enhance the rule-based paradigm. In *Proc. of the Second International Symposium on AI*, 1989.

[6] M. Jaeger. Probabilistic Reasoning in Terminological Logics. In Pietro Torasso Jon Doyle, Erik Sandewall, editor, *Proc. of the 4th Int. Conf. on Principles of Knowledge Representation and Reasoning*, pages 305–316, Bonn, FRG, May 1994. Morgan Kaufmann.

[7] D. Koller, A. Levy, and A. Pfeffer. P-Classic: A tractable probabilistic description logic. In *Proc. of AAAI 97*, pages 390–397, Providence, Rhode Island, 1997.

[8] Robert R. Korfhage. *Information Storage And Retrieval.* Wiley Computer Publishing, USA, New York, USA, 1997.

[9] S.L. Lauritzen and D.J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society*, 50:157–224, 1988.

[10] Zhaoyu Li and Bruce D'Ambrosio. Efficient inference in bayes networks as a combinatorial optimization problem. *Internatiional Journal of Approximate Reasoning*, 11:1–158, 1994.

[11] J. Lunze. *Künstliche Intelligenz für Ingenieure*, volume 2, chapter „Darstellung und Verarbeitung unsicheren Wissens“, pages 27–85. „Technische Anwendungen“, Oldenburg, 1995.

[12] Joel D. Martin and Dorrit O. Billman. Acquiring and combining overlapping concepts. *Machine Learning*, 16:121–155, 1994.

[13] R. Möller, V. Haarslev, and B. Neumann. Semantics-based information retrieval. In *Proc. IT&KNOWS-98: International Conference on Information Technology and Knowledge Systems*, pages 49–61, 31. August- 4. September 1998.

[14] R. Möller, K. Hidde, R. Joswig, T. Mantay, and B. Neumann. Band: Benutzeradaptiver Netzinformationsdienst. Technical report, Labor für Künstliche Intelligenz, FB Informatik, Universität Hamburg und „Lavielle“ EDV Systemberatung, GmbH & Co., Hamburg, 1999.

[15] Peter F. Patel-Schneider and Bill Swartout. Description Logic Specification from KRSS Effort. Technical report, 1993.

[16] J. Pearl. Fusion, propagation, and structuring in belief networks. *Journal of Artificial Intelligence*, 29:241–288, 1986.

[17] J. Pearl. *Probabilistic Resoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, Inc., San Mateo, California, 1988.

[18] Claas P. Prien. Anwendung der Least Common Subsumer-Methode zur Bestimmung von Ähnlichkeiten beim Information-Retrieval mit Beschreibungslogiken. Master's thesis, Universität Hamburg, 1998.

[19] V. Raghavan, P. Bollmann, and G. Jung. A critical investigations of recall and precision as measures of retrieval system performance. *ACM Transactions on Information Systems*, 7:205–229, 1989.

[20] L. Alperin Resnick, A. Borgida, R. Brachman, C.L. Isbell, D.L. McGuinness, P. F.Patel-Schneider, and K.C. Zalondek. *CLASSIC Description and Reference Manual for the Common Lisp Implementation.* 1995.

[21] Gerard Salton and Michael J. McGill. *Introduction to Modern Information Retrieval*, chapter Retrieval Evaluation, pages 157–197. McGraw-Hill Book Company, 1983.