

# NPBibSearch

## Eine Ontologie NP-vollständiger Probleme und deren Anwendung in einer Bibliografiercherche

Uwe Krüger

Friedrich-Schiller-Universität Jena  
Institut für Informatik

# NPBibSearch

Eine Ontologie NP-vollständiger Probleme und deren Anwendung in der Bibliografiercherche

## Inhalt

1. Die Komplexitätsklasse NP
2. Ontologie NP-vollständiger Probleme
3. NPBibSearch
4. Zusammenfassung

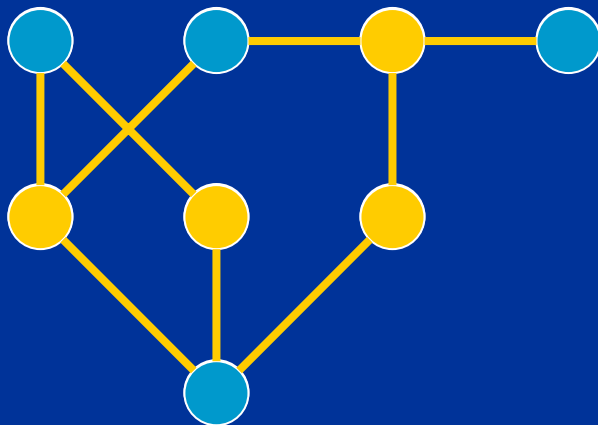
# 1. Die Komplexitätsklasse NP

- Es gibt viele praxisrelevante Probleme für die sich bis heute **keine** effizienten Algorithmen finden lassen.
- Gemeint sind sog. **Entscheidungsprobleme**, die so formuliert sind, dass ein möglicher Lösungs-Algorithmus entweder die Antwort **ja** oder **nein** liefert. (0/1)
- Alle bekannten Algorithmen für diese Probleme erfordern einen **exponentiellen** Rechenaufwand.
- In der Komplexitätstheorie werden diese Probleme unter der Klasse **NP** zusammengefasst. (Non-deterministic Polynomial time)
- Z.B. **Satisfiability**, **Vertex Cover**, **Clique**, etc...

# 1. Die Komplexitätsklasse NP

- Beispiel **Vertex Cover**

- Gegeben: Ein ungerichteter Graph  $G = (V, E)$ ; und eine natürlicher Zahl  $k$ .
- Frage: Gibt es eine Teilmenge  $C \subseteq V$  mit  $|C| \leq k$ , sodass jede Kante aus  $E$  min. einen Endpunkt in  $C$  hat?



Vertex Cover für  $k = 4$  ●

Bsp. **Kombinatorische Explosion**

n / k	Laufzeit für $O(n^k)$
50 / 4	0.06 sec
150 / 10	$1.8 \times 10^6$ Jahre

(Annahme:  $10^8$  Op./sec)

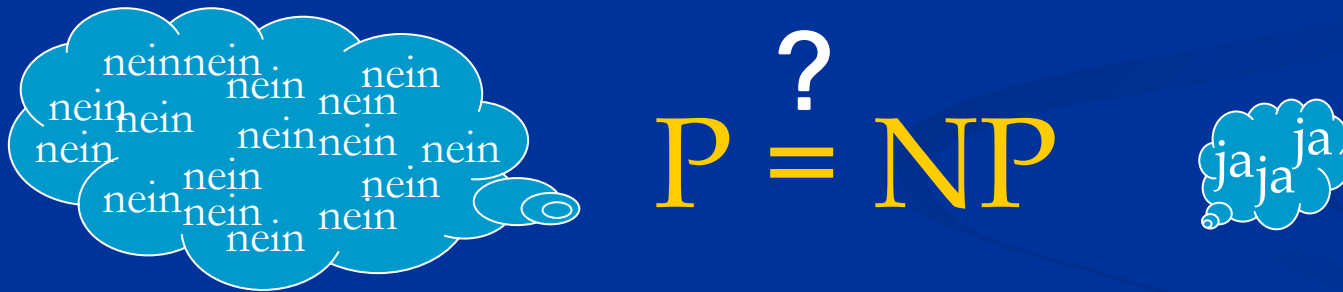
# 1. Die Komplexitätsklasse NP

Informelle Begriffsdefinitionen:

- **Reduktion:**  
Eine **Reduktion** ist eine Transformation, die eine Probleminstance von einem Problem  $L'$  auf eine Probleminstance eines anderen Problems  $L$  mit **polynomiellen** Berechnungsaufwand transformiert.
- **NP-schwer:**  
Ein Entscheidungsproblem  $L$  heißt **NP-schwer** (NP-hard) genau dann, wenn: alle Probleme aus der Klasse **NP** **polynomiell** auf  $L$  **reduzierbar** sind.
- **NP-vollständig:**  
Ein Entscheidungsproblem  $L$  heißt **NP-vollständig** (NP-complete) genau dann, wenn:  $L$  in der Klasse **NP** liegt und  $L$  **NP-schwer** ist.

# 1. Die Komplexitätsklasse NP

- Sind NP-vollständige Probleme **effizient** lösbar?
- Viele Probleme aus NP, insbesondere die NP-vollständigen, lassen sich **vermutlich** nicht effizient lösen.
- Hieraus resultiert eine der großen heute noch **offenen** Fragestellungen der theoretischen Informatik:



- Die Fragestellung gehört zu den sog. **Millennium Prize Problems** des **Clay Mathematics Institute**, für deren Lösung eine Million \$ Preisgeld ausgesetzt wurde.

# NPBibSearch

Eine Ontologie NP-vollständiger Probleme und deren Anwendung in der Bibliografierforschung

## Inhalt

1. Die Komplexitätsklasse NP
2. **Ontologie NP-vollständiger Probleme**
3. NPBibSearch
4. Zusammenfassung

# 2. Ontologie NP-vollständiger Probleme

## Technische Umsetzung

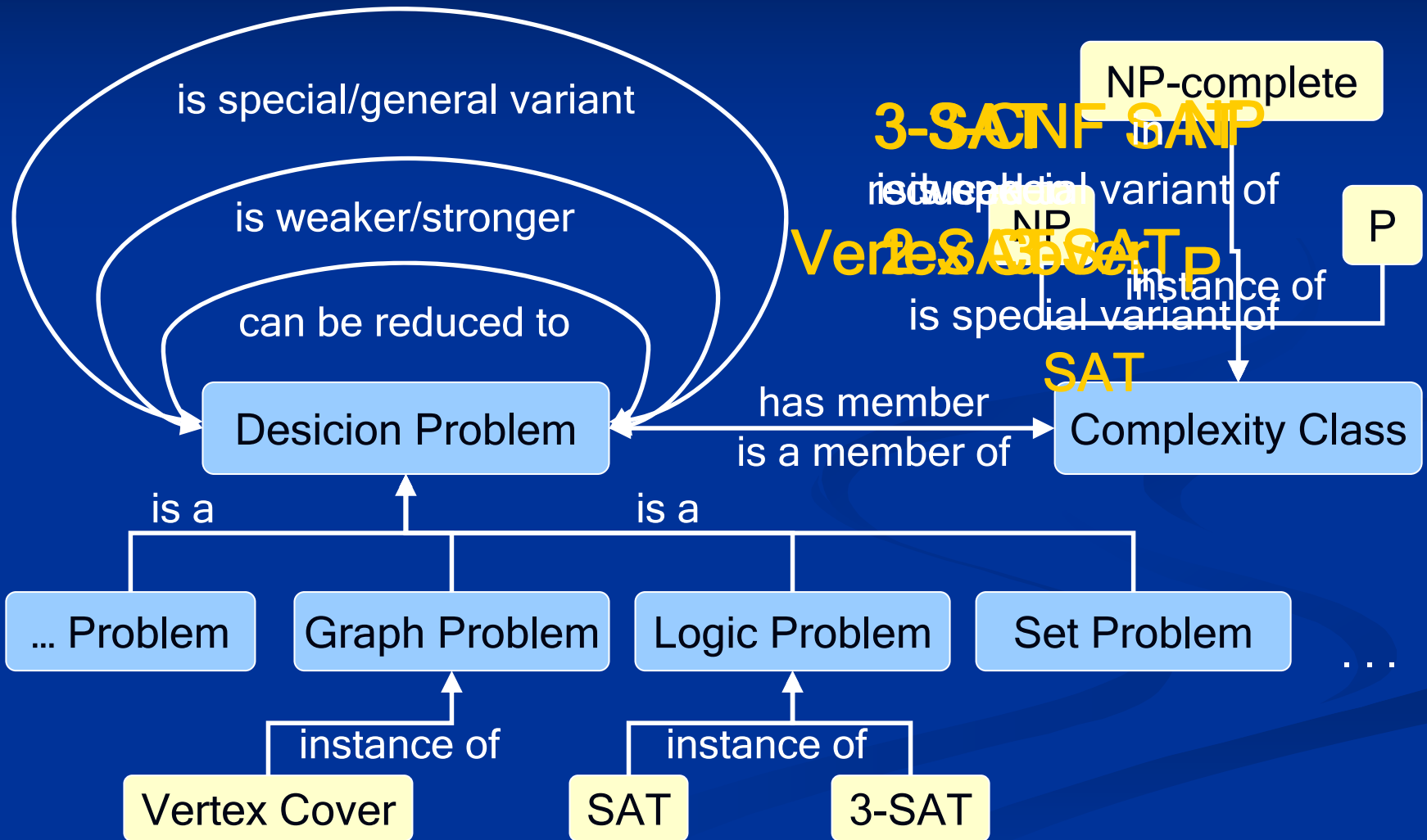
- Ontologie Engineering? Welcher Ansatz?
- Einsatz von **Protégé-OWL** (Version 3.2)
- Sprachebene: **OWL LITE**, DL oder FULL?
- T-Box und A-Box in einer OWL-Datei
- **T-Box:**
  - **Classes** und **Properties** für Konzept der NP-vollständigen Entscheidungsprobleme
- **A-Box:**
  - Standardwerk von **Garey & Johnson** 1979

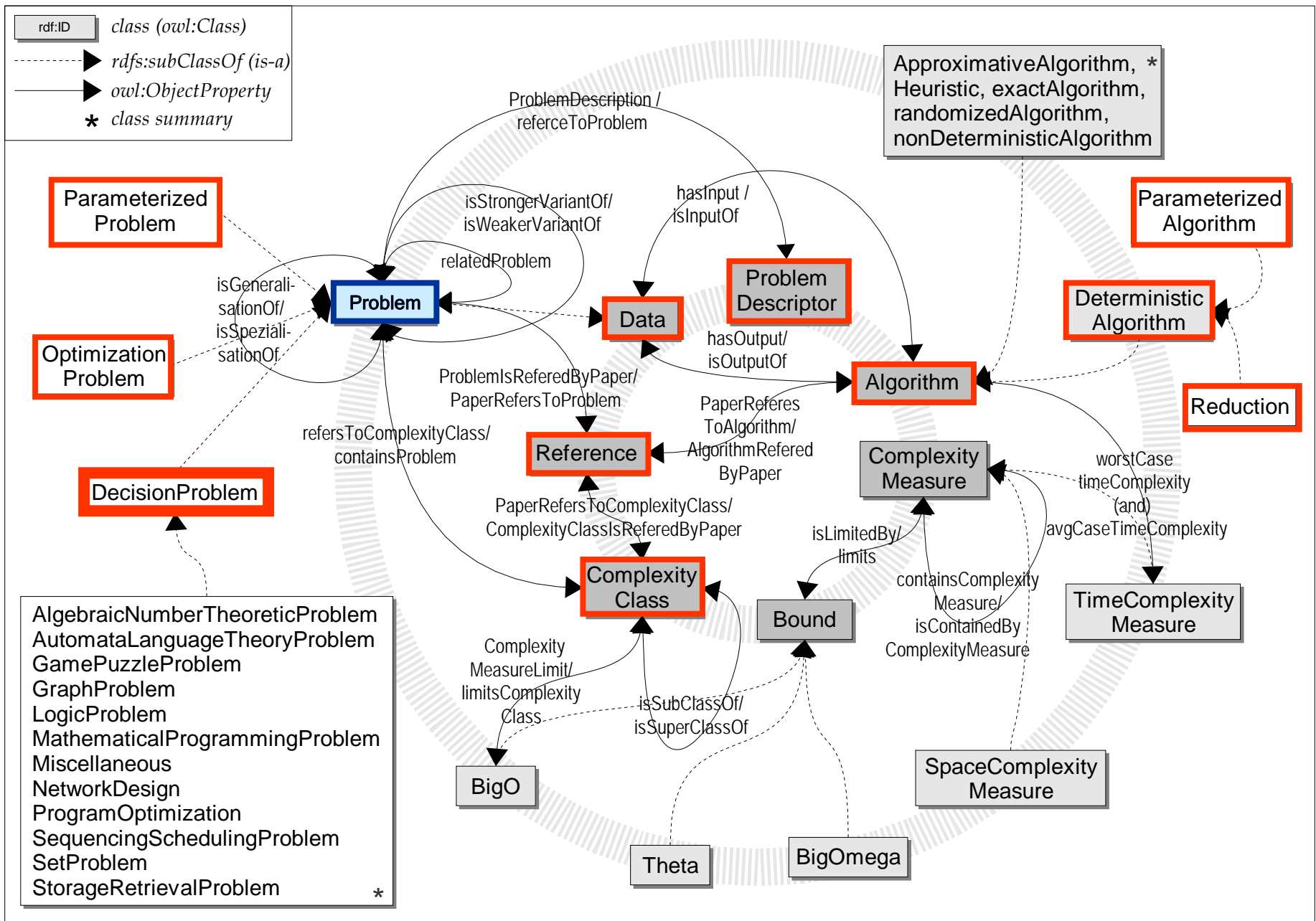
## 2. Ontologie NP-vollständiger Probleme

### Ontology Engineering - Zyklus:

1. Ermitteln der **Basisbegriffe** und Konzepte der gewünschten Domainmodellierung
2. Definieren der **Klassenhierarchie** (Class/subClass)
3. Definieren der **Eigenschaften** (Object/Datatype-Properties)
4. Festlegen: Wertebereich, Kardinalitäten, Beziehungstypen (symmetrisch, transitiv, etc.)
5. **Instanzen Definition** (Eingabe der Individuen)
6. **Evaluierung**: Syntax-Check, Überprüfung auf mögliche Inkonsistenzen, etc.  
(Ontologie überarbeiten/anpassen)

# 2. Ontologie NP-vollständiger Probleme: Domain-Modellierung

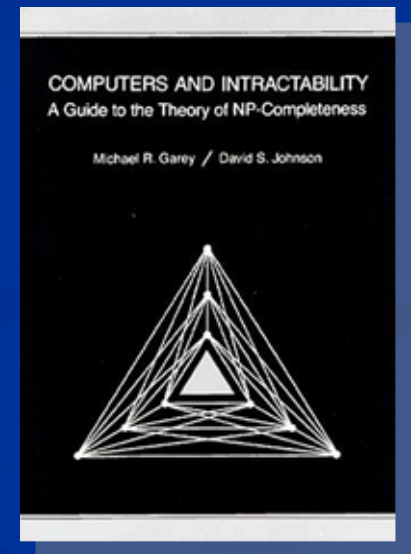




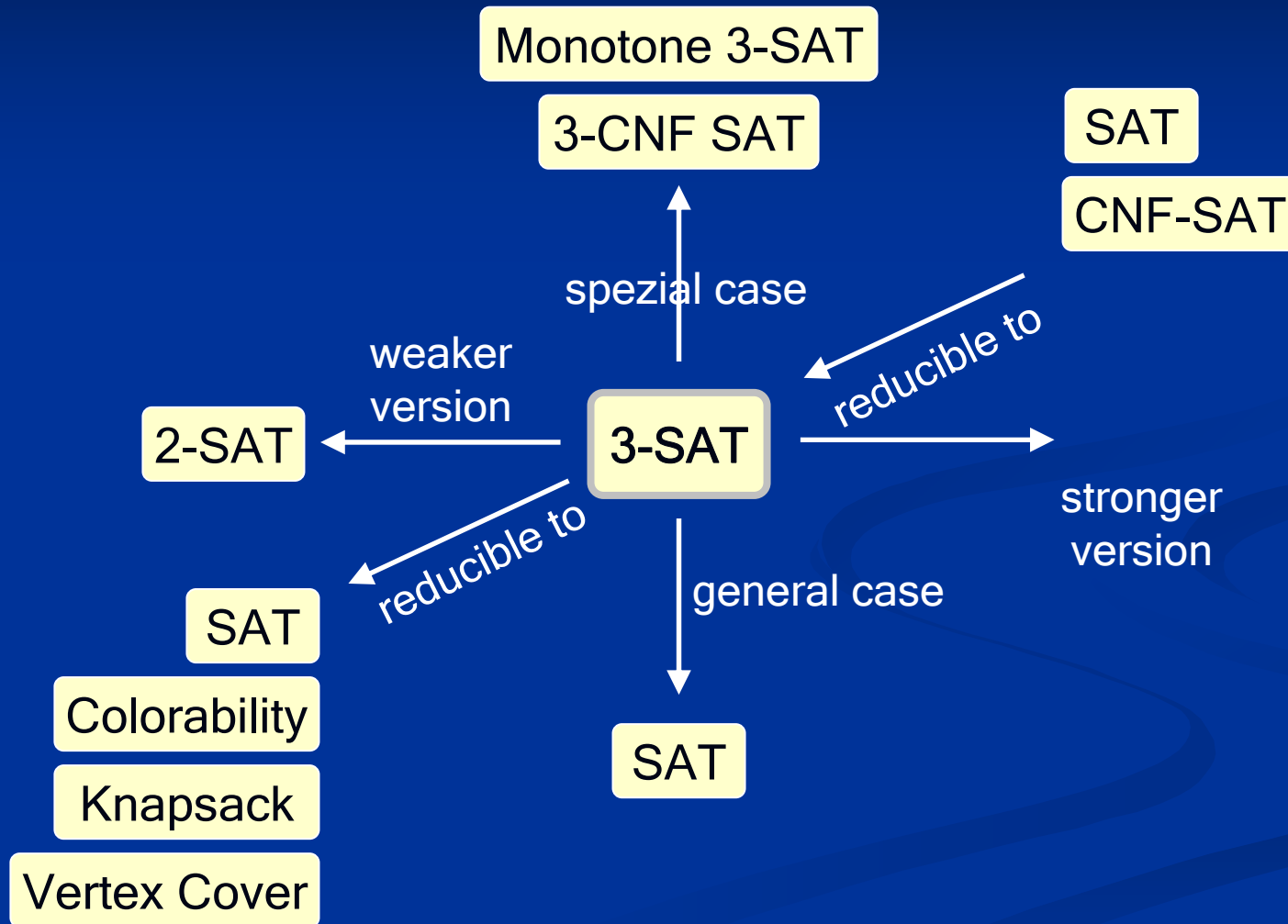
## 2. Ontologie NP-vollständiger Probleme

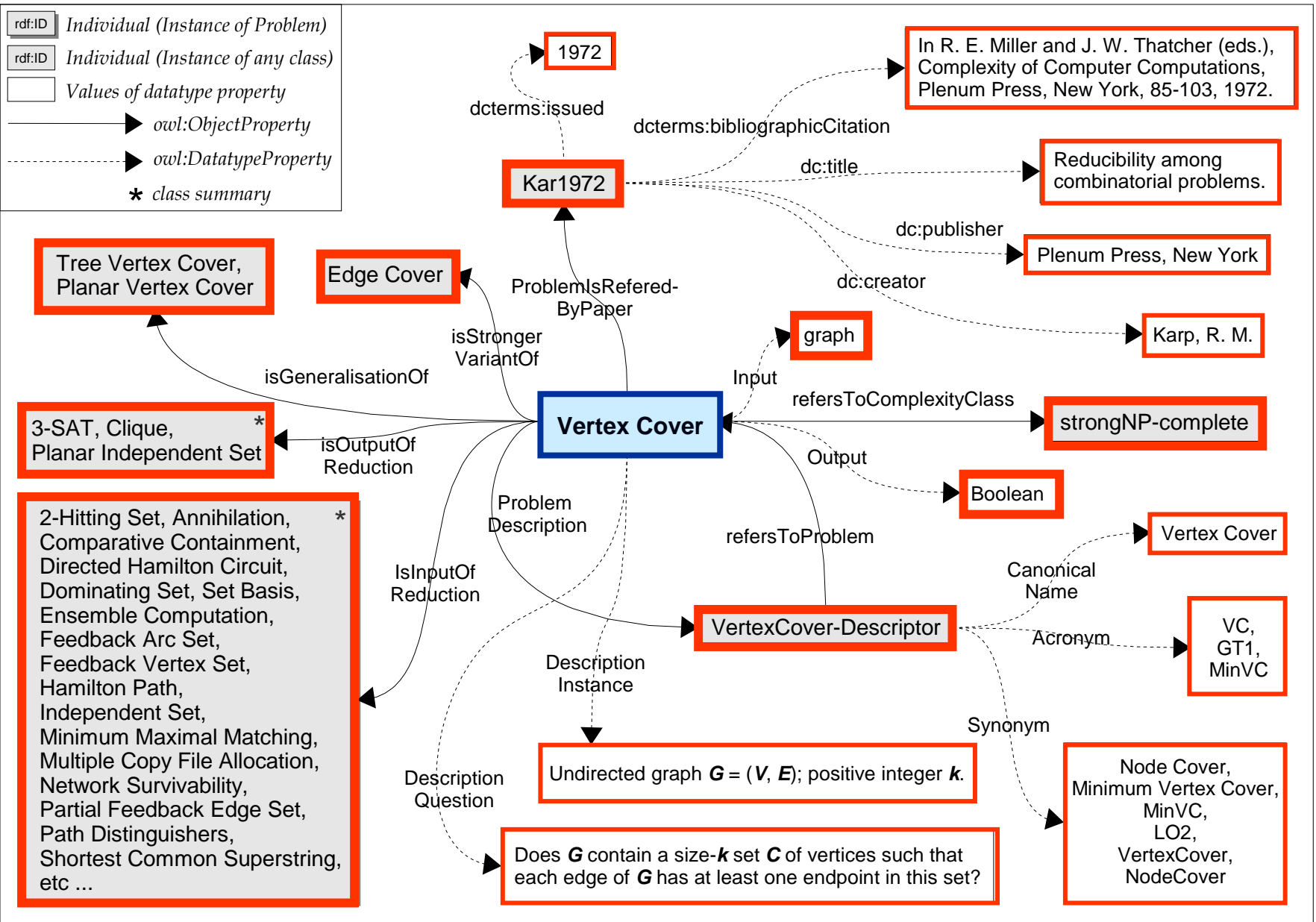
### Anlegen der Klasseninstanzen (**A-Box**)

- Problemsammlung aus dem Buch von **Garey und Johnson** 1979: „Computers and Intractability: A Guide to the Theory of NP-Completeness“
- ca. **350** NP-vollständige Probleme (mit Reduktionen)
- Angabe von **Beziehungen** zwischen den Problemen die über die Reduktion hinaus gehen
- Problem**beschreibungen**: Problemname, Synonyme, Instance, Question, References, Notes, etc...



## 2. Ontologie NP-vollständiger Probleme





# NPBibSearch

Eine Ontologie NP-vollständiger Probleme und deren Anwendung in der Bibliografiercherche

## Inhalt

1. Die Komplexitätsklasse NP
2. Ontologie NP-vollständiger Probleme
3. **NPBibSearch**
4. Zusammenfassung

# 3. NPBibSearch

## Web-Suche und Ontologien

Wie können Ontologien für eine Bibliografie-Recherche genutzt werden?

1. **Suchstring Evaluierung**  
Wonach genau sucht der Nutzer? Was ist gemeint?  
(query string matching)
2. **Suchstring Erweiterung**  
Suchraumerweiterung oder -verkleinerung. Modellierung eines angepassten Suchstrings durch AND/OR Verknüpfungen.
3. **Domain-Navigation** und Querverweise  
Dem Nutzer bei seiner Suche unterstützen: Querverweisen aus der Taxonomie und ObjectProperties dafür nutzen.
4. Bereitstellen **zusätzlicher Informationen**  
Extrahieren möglichst vieler relevanter Fakten bezüglich der Suchanfrage.

# 3. NPbibSearch: ECCC - The Electronic Colloquium on Computational Complexity

- Elektronisches Wissenschaftsjournal für Inhalte aus dem Gebiet der **Computational Complexity**
- **Wissenschaftliche Paper** (Sammlung von elekt. Büchern, Diplom-, Dissertationsarbeiten und Vorlesungsmaterialien.)
- **Ziel:** Neue wiss. Ergebnisse schnell bereitstellen
- Seit **1994**, Prof. Dr. Christoph **Meinel**
- <http://eccc.hpi-web.de/eccc/> >**1000** Arbeiten



# 3. NPBibSearch - Allgemein

- NPBibSearch ist eine Web-basierte **ontologie-gestützte** bibliografische **Suchmaschine** für die online ECCC-Datenbank
- **Volltextsuche** auf ECCC analog zu allg. Websuchmaschinen
- Ontologieunterstützung ist beschränkt auf die Domain der NP-vollständigen Probleme.
- Dem Nutzer wird zusätzlich zur Volltextsuche eine (einfache) Ontologie-basierte **Navigation** durch den Suchraum (Domain) geboten.

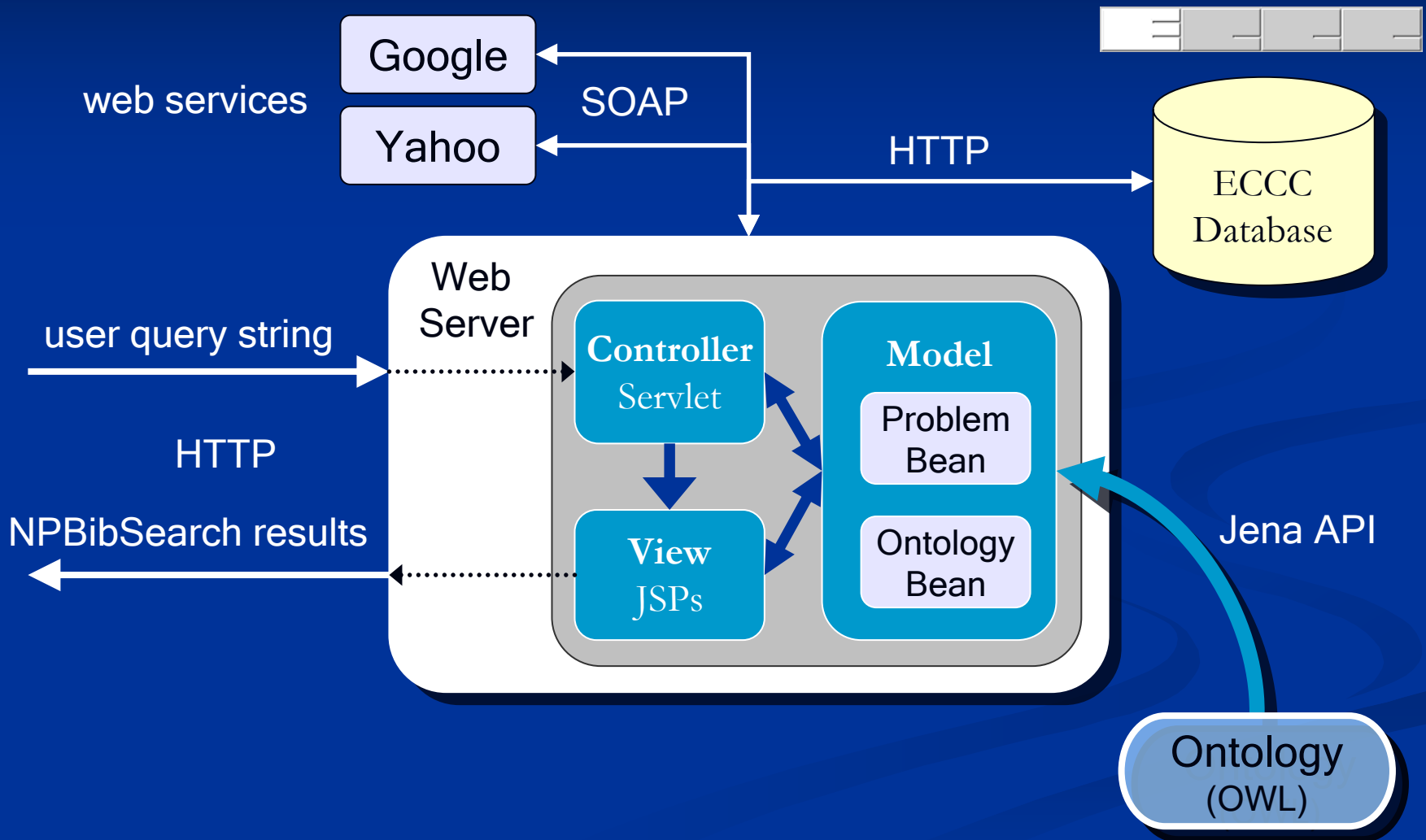


# 3. NPBibSearch - Technik

- Eine auf **Java-Servlets** basierende Webanwendung zur bibliografischen Suche in der ECCC-Datenbank
- Ontologieunterstützung erfolgt auf **vier** Wegen:
  - Suchstring **Validierung**
  - Suchstring **Erweiterung**
  - **Domain-Navigation** mittels Querverweisen
  - **Zusatzinformation** (Beschreibung, Referenzen und Synonyme) (Suche nach konkreten Entscheidungsproblemen)
- Nutzung der Volltextsuche über **Google** oder **Yahoo**.



# 3. NPBibSearch - Architektur



NP BibSearch

Yahoo
  Google

Search

3-sat

You can choose synonyms to expand your search:

LO2
  3-CNF-SAT
  3 SAT
  3SATISFIABILITY

3-SAT
  3SAT
  3-SATISFIABILITY

hide synonyms

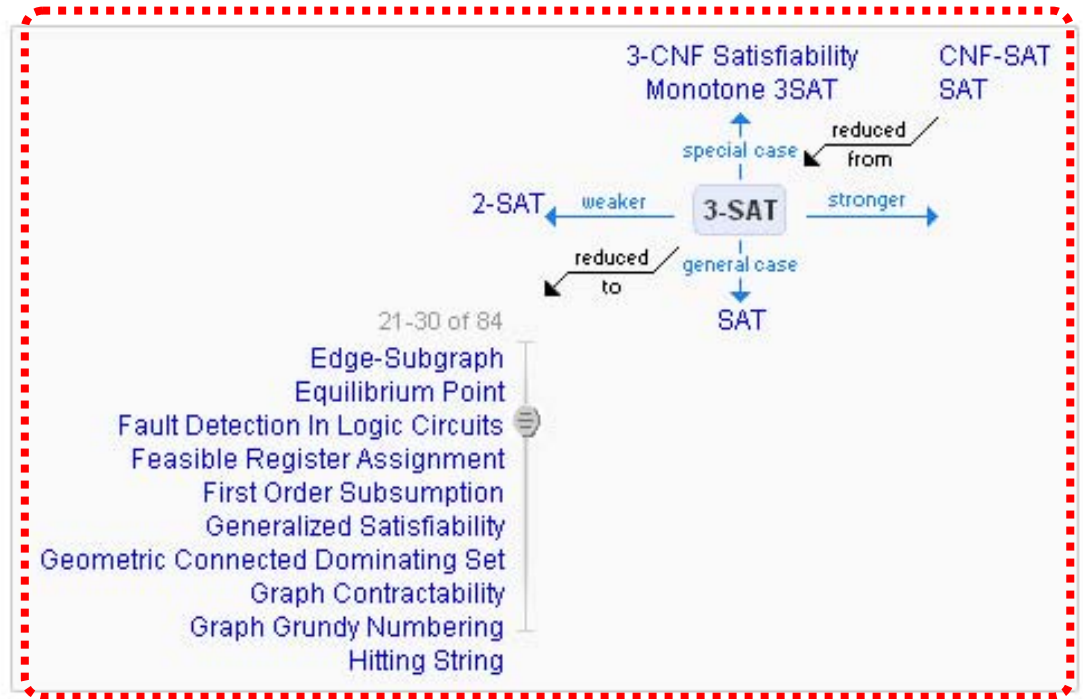
**3-SAT**

**Instance:** Set  $U$  of variables, collection  $C$  of clauses over  $U$  such that each clause  $c \in C$  has  $|c|=3$ .

**Question:** Is there a satisfying truth assignment for  $C$ ?

**Reference:** [Coo1971a]

Cook, S. A., *The Complexity of Theorem-Proving Procedures*, Proc. 3rd Ann. ACM Symp. on Theory of Computing, Association for Computing Machinery, New York, 151-158, 1971.



ECCC (Google) Results 1-10 of about 50 from [eccc.lpi-web.de](http://eccc.lpi-web.de) for 3-sat.

[around the document](#)
[keywords](#)
[third party links](#)

1. ECCC Report TR03-053, accepted on Jul 08, 2003 | BibTeX

[KAZUO IWAMA](#)
[SUGURU TAMAKI](#)

Improved Upper Bounds for **3-SAT** (PS | PDF)

This paper presents a new upper bound for the  $k$ -satisfiability problem. For small  $k$ 's, especially for  $k=3$ , there have been a lot of algorithms which run significantly faster than the trivial  $2^n$  bound. The following list summarizes those algorithms w... + more

(Google snippet)

CNF Satisfiability, Probabilistic Algorithm, Complexity,

[CiteSeer](#)
[dblp over Google](#)
[CSBC](#)

2. ECCC Report TR03-007, accepted on Jan 28, 2003 | BibTeX

[OLIVIER DUBOIS](#)
[YACINE BOUFGHAD](#)
[JACQUES MANDLER](#)

Typical random **3-SAT** formulae and the satisfiability threshold (PS | PDF)

$k$ -SAT is one of the best known among a wide class of random constraint satisfaction problems believed to exhibit a threshold phenomenon where the control parameter is the ratio, number of constraints to number of variables. There has been a large am... + more (Google snippet)

satisfiability, random 3-SAT, threshold phenomena, phase transitions,

[CiteSeer](#)
[dblp over Google](#)
[CSBC](#)

# NPBibSearch

Eine Ontologie NP-vollständiger Probleme und deren Anwendung in der Bibliografiercherche

## Inhalt

1. Die Komplexitätsklasse NP
2. Ontologie NP-vollständiger Probleme
3. NPBibSearch
4. **Zusammenfassung**

# 4. Zusammenfassung

Was wurde erreicht:

- + NPbibSearch nutzt die Ontologie als Background-Wissen für erweiterte Suchfunktionalitäten.
- + Der Nutzer kann sich bei seiner Recherche durch den „Domainraum“ führen lassen.
- + Dem Nutzer werden Querverweise auf andere Probleme offeriert, die für ihn bis dahin evt. unbekannt waren.
- Qualitätssicherung der Ontologie ist aufwendig und noch nicht abgeschlossen. (Review von Fachleuten notwendig)

# 4. Zusammenfassung

## Weiterführende Arbeit:

- Erweiterung der Ontologie durch das Konzept der **Parametrisierten Probleme**.
- Notwendigkeit einer universellen Benutzerschnittstelle zur gemeinsamen Erweiterung und **Pflege der Ontologie**.
- Ausweitung der bibliografischen Recherche auf **andere online Bibliografie-Datenbanken**.
- **Aufweitung der Domain** um weitere geeignete Wissensbereiche der Informatik.
- Kollaborative Verschlagwortung von Publikationen (**Tagging**).

# Vielen Dank für Ihre Aufmerksamkeit !

<http://ipc755.inf-nf.uni-jena.de:8084/NPBibSearch/>  
oder über  
[www.minet.uni-jena.de/~sack](http://www.minet.uni-jena.de/~sack)