

und ermöglichen eine weitere Navigation im Datenbestand. Vom Autor einer Publikation gelangt man per Mausklick auf dessen individuelle Publikationsliste, von einem Zeitschriftentitel zum Inhaltsverzeichnis oder zum Jahresregister der Zeitschrift. Mehrfachtreffer im Suchergebnis werden durch ein Werkzeug weitestgehend ermittelt und in einem einheitlichen Datensatz zusammengeführt. Die Verweise auf die Originalquellen, die unterschiedlichen Umfang haben, werden angezeigt. Ausgehend von den Suchergebnissen wird eine Volltextvermittlung oder der direkte Zugriff auf elektronische Volltexte angeboten, z.B. auf Tagungsberichte, Dissertationen und Zeitschriften verschiedener GI-Fachgruppen sowie weiterer Anbieter (u.a. Verlage, Fachgesellschaften, Hochschulen). Der direkte Zugriff auf die elektronischen Volltexte der LNI-Reihe (Lecture Notes in Informatics) der GI wird exklusiv über *io-port.net* angeboten.

Für ausgewählte Fachbereiche der Informatik wurden exemplarisch Themenseiten eingerichtet, die einen kompakten Überblick über aktuelle Forschungs- und Entwicklungsergebnisse, laufende Projekte, Publikationen, sowie relevante Links und aktuelle Veranstaltungen geben. Eine redaktionelle Betreuung dieser Themenseiten ist Voraussetzung für deren Erfolg. Als weitere Komponente von *io-port.net* wird ein Werkzeug angeboten, das die ortsunabhängige Erstellung und Verwaltung von persönlichen bzw. institutsbezogenen Publikationslisten ermöglicht. Weitere Fachinformationen in *io-port.net* sind: Informatik-Lexikon, allgemeine Linklisten sowie ein umfassender Konferenzkalender.

Semantische Werkzeuge und personalisierte Dienste unterstützen eine effiziente Recherche. *io-port.net* dient als Anwendungsumgebung für die in dem forschungsorientierten Projekt „Semantische Methoden und Tools für Informationsportale (SemIPort)“ neu entwickelten Werkzeuge und integriert diese in die wissenschaftliche Informationsversorgung.

1.2 Suche nach Fachinformation im Zeitalter von Google Scholar

Seit November 2004 bietet Google unter scholar.google.com einen speziellen Suchdienst für wissenschaftliche Recherchen an. Google scholar durchsucht nach Angaben der Betreiber Bücher, wissenschaftliche Abhandlungen, technische Dokumente, Fachzeitschriften und sonstige Literatur aus dem Umfeld von Forschung und Lehre. Einen bedeutenden Vorteil gegenüber Google [Cu05] erlangt dieser Suchdienst durch die Verwendung einer Zitationsanalyse für Publikationen (Cited by). Allerdings ist der Suchalgorithmus noch nicht effizient, weil Suchergebnisse nicht immer zu Volltexten führen und damit wenig nutzbringend sind. Außerdem sind viele Suchergebnisse nicht anzeigbar, weil es sich um Zitate handelt. Das Informationsportal *io-port.net* hat sich mit seinen Informationsangeboten und integrierten Suchfunktionalitäten im Gegensatz zu Google scholar auf ein Fachgebiet spezialisiert - die Informatik - und deckt dieses Gebiet umfassend ab [Ko04]. In *io-port.net* sind auch semantische Werkzeuge integriert, die den Benutzer bei seiner Suche nach relevanten Inhalten unterstützen. Die Literaturnachweise sind aufgrund vollständiger Angaben und Abstracts qualitativ hochwertig und nach der Anzeige von Suchergebnissen gelangen Nutzer zur Volltextvermittlung. Gegenüber Google scholar finden Nutzer in *io-port.net* Dokumentationen zur Nutzung der Suchfunktionalitäten und zu den semantischen Werkzeugen. Ein weiterer Vorteil von *io-*

port.net besteht in der Klärung rechtlicher Fragestellungen zum Volltextdownload oder zur Referenz auf Volltexte.

In einem Kurzttest ergab die Suche in Google scholar nach dem Begriff „evolutionäre Algorithmen“ 538 Treffer, wobei auf den ersten beiden Seiten schon 15 Treffer nicht angesehen werden konnten – die Suchmaschine hatte die Texte nur als Zitate in anderen wissenschaftlichen Werken gefunden. In *io-port.net* ergab die Suche nach diesem Begriff 65 Treffer. In Google scholar werden zwar mehr Suchtreffer angezeigt, allerdings handelt es sich um viele Zitate oder Buchtreffer und erfahrungsgemäss beachten Suchmaschinennutzer auch nur die ersten Treffer einer Anzeige. Die englischsprachige Suche nach „evolutionary algorithms“ ergab eine Trefferzahl von 183.000, hier sind unter den ersten 100 Treffern 21 Zitate und 10 Büchertreffer. In *io-port.net* werden 4617 Treffer an wissenschaftlichen Publikationen und Zeitschriftenartikeln angezeigt. Ähnliche Unterschiede in der Treffergröße zwischen Deutsch und Englisch gibt es, wenn nach dem Begriff „Entwurfsmuster“ beziehungsweise „design pattern“ gesucht wird. Der englische Begriff liefert 831.000 Treffer und der deutsche 941. In *io-port.net* ergeben sich ähnliche Unterschiede bei der Trefferanzahl zwischen Englisch und Deutsch: für den englischen Begriff können 3555 und den deutschen 39 Treffer angesehen werden. Damit Nutzer in *io-port.net* bei hohen Trefferzahlen relevante Inhalte finden, können durch die Integration von semantischen Werkzeugen verfeinerte Begriffe der Suche angezeigt werden: bei der Suche nach dem Begriff „design patterns“ werden noch Begriffe wie „design patterns learning“, „design patterns problem“, „design patterns applications“, „system design patterns“ oder „software design patterns“ vorgeschlagen.

Sowohl Google scholar als auch *io-port.net* befinden sich momentan in der Test- bzw. Anfangsphase und müssen sich gegenüber verschiedenen Herausforderungen (u.a. Dublettenerkennung, Extraktion der Autorennamen, Integration von Volltexten) positionieren. Als Fazit des Kurzttests lässt sich feststellen, dass der Betastatus Google scholar anzumerken ist und die Ergebnisse mancher Suchbegriffe lückenhaft und nicht aktuell sind. Die Recherche in einem Informationsportal wie *io-port.net* kann Google scholar auf keinen Fall ersetzen.

1.3 Erfassung und Bearbeitung bibliographischer Daten

Die klassische Erfassung bibliographischer Daten erfordert viel Aufwand und Zeit. Mit geeigneten Methoden können diese Daten aus Inhaltsverzeichnissen im HTML-Format halbautomatisch extrahiert und aufbereitet werden. Werkzeuge unterstützen die Erfassung und Korrektur der gewonnenen Daten und ermöglichen so eine schnelle und korrekte Erfassung bibliographischer Daten.

Die Anwendung **DataGen** extrahiert aus Inhaltsverzeichnissen, die im HTML-Format vorliegen, bibliographische Daten mit Hilfe von Wrappern. Vordefinierte Masken ergänzen die gewonnenen Daten um weitere Inhalte. DataGen ist in der Lage, weitere Daten aus dem Internet nachzuladen, zu verarbeiten und mit den bisher gewonnenen Daten zu vereinigen. Mit nachgeschalteten Skripten erfolgt eine automatische Aufbereitung und