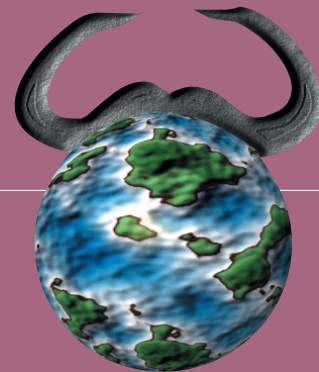


Die monatliche GNU-Kolumne

Brave GNU World



Diese Kolumne berichtet aus der Perspektive des GNU-Projekts und der FSF über Projekte und aktuelle Geschehnisse aus dem Umfeld freier Software. In dieser Ausgabe geht es um die Projekte Doodle und Pylint sowie um Privacy-Probleme beim Online-Portal Orkut. Georg C. F. Greve

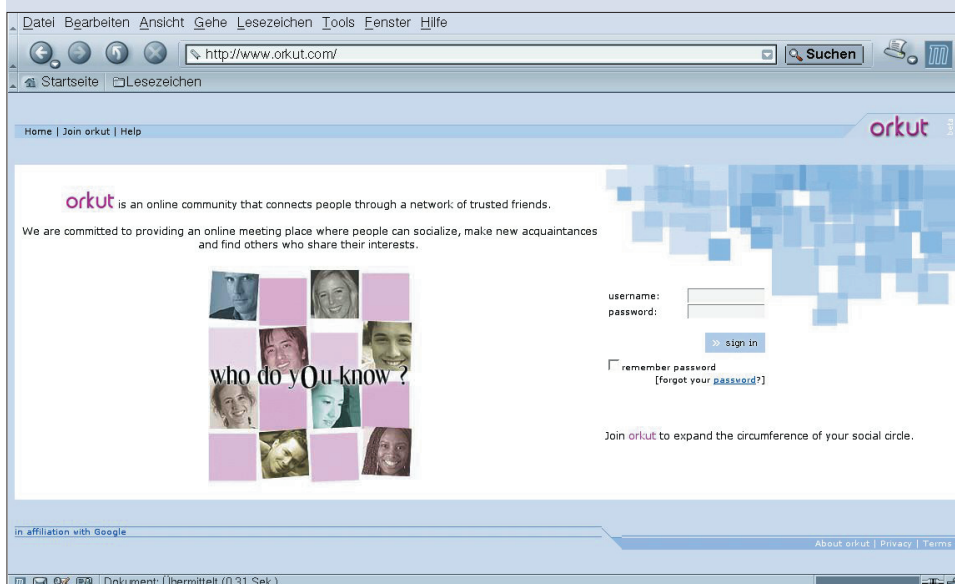


Abbildung 1: Das Online-Portal Orkut verspricht seinen Teilnehmern eine gezielte Suche nach Gleichgesinnten durch ein umfassendes Persönlichkeitsprofil. Aber wie sieht es mit dem Schutz der sensiblen Daten aus?

Willkommen zu einer neuen Ausgabe der Brave GNU World. In diesem Monat sollen zwei den meisten Lesern vermutlich unbekannte Projekte zu ihrem Recht auf Öffentlichkeit kommen. Am Beginn steht jedoch ein Privacy-Problem, das sich aus der populären Community-Website Orkut ergibt.

Orkut

Wer hat schon mal von Orkut gehört oder besitzt dort einen Account? Die Site [1] ist eine von Google [2] angebotene Plattform zur Bildung und Pflege von sozialen Netzwerken. Geschrieben wurde sie von Orkut Buyukkokten, einem Google-Angestellten, dessen Namen sie trägt. Angestellte der Firma Google sind angehalten, etwa ein Fünftel ihrer Arbeitszeit in persönliche Interessen zu stecken, und Buyukkokten interessierte sich für soziale Netzwerke.

Voraussetzung für die Teilnahme am Netzwerk ist eine Einladung: Ein bereits eingetragener Benutzer gibt an, dass er eine bestimmte Person als Freund in sein Netzwerk aufnehmen möchte, und diese Person bekommt dann per E-Mail eine Einladung zugeschickt.

Soziale Netzwerke

Im ersten Schritt legt der neue Benutzer ein Profil an, in dem neben Geburtstag, Geschlecht, Wohnort, Interessen, Humor, sexueller Orientierung und aktuellem Beziehungsstatus noch andere Dinge Platz finden. Allerdings sind alle Angaben freiwillig, die über den Namen und ein Login hinausgehen. Danach zieht der neu eingetragene Nutzer los und sucht nach Bekannten und Freunden, um diese wiederum in sein Orkut-Netzwerk aufzunehmen. Neben dem eigenen Namen blendet die Seite

immer die Anzahl der Freunde ein. Ein Anreiz, möglichst viele Freunde und Bekannte zu vernetzen.

Zusätzlich zu dieser einfachen Kontaktaufnahme bietet die Site die Möglichkeit, andere Teilnehmer zu bewerten. So steht es jedem frei, eines der drei Kriterien Trusty, Cool oder Sexy zu vergeben oder sich gar zum Fan einer bestimmten Person zu erklären. Themen- und Interessengruppen runden das soziale Networking ab. Entweder schließt sich der Nutzer selber einer Gruppe an oder er lädt andere in eine solche Gruppe ein, um mit ihnen zu diskutieren.

Mit anderen Worten: Orkut besitzt ein umfassendes Persönlichkeitsprofil jeder eingetragenen Person, das sehr individuelle und private Bereiche von politischer Einstellung bis hin zu sexuellen Neigungen umfassen kann.

Datenschutz bedroht

Dieses Netz aus sozialen und thematischen Informationen, das Orkut zu den einzelnen Personen sammelt, wirkt sich bedrohlich auf den Datenschutz aus. Die Frage, was mit den gespeicherten Daten geschieht, beantwortet das Privacy Statement auf der Webseite – doch nicht so umfassend, wie es wünschenswert wäre. Die Betreiber der Site informieren darüber, dass im „Interesse des Service und der Erkennung von technischen Problemen“ die folgenden Informationen bei jedem Besuch erhoben werden: verwendeter Browser, eingestellte Sprache, IP-Adresse und Statistiken über die Nutzung der Seite im Bezug darauf, was wie oft angesehen wird.

Ein weiterer Pferdefuß: Mit der Nutzung von Orkut stimmt der Anwender der

Speicherung und Nutzung aller Daten in den USA sowie in jedem anderen Land zu, in dem Orkut oder Google Abteilungen betreiben. Für den Fall, dass die Firma Orkut aufgekauft wird oder mit einem anderen Unternehmen fusioniert, erhält Orkut sogar das Recht, die gesammelten Daten weiterzugeben. Sollte dieser Fall jedoch eintreten, wollen die Betreiber die Anwender zumindest vorab informieren.

Beklemmendes Bild

Und noch etwas anderes gibt das Unternehmen ebenfalls nicht preis: Sollte ein US-Gericht es verlangen, sind der Zugriff auf Daten sowie die Speicherung der gesammelten Daten des persönlichen Accounts ausdrücklich gestattet. Die Speicherung ist schon deshalb gestattet, um auf einen möglichen juristischen Prozess vorbereitet zu sein.

Es ist zudem nicht vorgesehen, den Anwender zu informieren, wenn Orkut seine Daten an Google weitergibt, was nach der Privacy Policy eindeutig erlaubt ist. Google wiederum speichert seit Anbeginn die IP-Adresse zu jeder Suchanfrage. In Anbetracht der ausgiebigen Verwendung von Cookies, ohne die die Seite nicht vernünftig funktioniert, ergibt sich ein beklemmendes Bild.

Dieser Komplex aus Suchmaschine, dem E-Mail-Dienst Gmail [3] sowie dem so-

zialen Netzwerkforum Orkut lässt mit Sicherheit bei zahlreichen mündigen Bürgern und Datenschutzbeauftragten die Alarmglocken läuten. Die jüngste Auseinandersetzung über die Weitergabe von Passagierdaten bei Flügen von Europa in die USA erscheint tatsächlich wenig aufregend angesichts des Gefahrenpotenzials dieser Sammel- und Verwertungswut eines einzigen US-Unternehmens.

Daraus sollten die Benutzer Konsequenzen ziehen. Wer nicht bei Orkut eingetragen ist, sollte dies tunlichst vermeiden. Wer bereits eingetragen ist, kann austreten. Das hilft allerdings nicht viel, denn in der Privacy Policy steht ausdrücklich, dass die Terminierung des Accounts dazu führt, dass das Profil und die weiteren Informationen „nicht mehr sichtbar sind.“ Die Firma löscht die Daten also wahrscheinlich nicht. Ein denkbarer, wenngleich wenig effektiver Ausweg: die Angabe falscher Daten. Hier besteht die Gefahr, dass die Betreiber aus der Art der Fälschungen ebenso viel über den User erfahren.

Ersetzt Orkut!

Zwei Dinge überraschen an Orkut besonders: Die augenscheinlich große Zahl von technisch versierten Menschen, die völlig unkritisch an die Sache herangehen und sorglos ihre Daten eintragen.

Und zum anderen der große Erfolg – es gibt anscheinend tatsächlich Bedarf. Wenn dies den Tatsachen entspricht, sollte sich die Community darum bemühen, George Orwell aus Orkut zu entfernen. Wir brauchen eine Implementation mit der gleichen Funktionalität, die jedoch die Privatsphäre der Menschen achtet und schützt.

Ein paar Gedanken dazu: Wenn die Daten auf dem Server liegen, gibt es Begehrlichkeiten bezüglich der Verwendung, für welchen Zweck auch immer. Idealerweise bleibt das komplette Profil eines Nutzers also in der Hand des Anwenders und der Server speichert nur wenige Informationen. Sensible Daten erhält er nur bei Bedarf vom Client und löscht diese auf Wunsch wieder.

Kommt bei Client und Server freie Software zum Zuge, dann ist zusätzlich Transparenz gewährleistet, weil der Nutzer die Kontrolle über die Hülle erhält, die seine Daten umschließt. Außerdem ermöglicht der Einsatz von Verschlüsselung eine sichere Übertragung der Daten vom Client zum Server sowie eine sichere Speicherung auf beiden Seiten. Ein gutes Beispiel für den erfolgreichen Einsatz einer solchen Kombination aus verschiedenen Techniken ist das Freenet-Projekt [4]. Hier wissen die Rechner beispielsweise nicht, welche Information auf ihnen abgelegt ist.

Da eine freie Implementation von Orkut keine Hintertüren auflassen sollte, muss die technische Debatte von ausgewiesenen Fachleuten auf diesem Gebiet geführt werden. Daher der Aufruf an alle, die so etwas tun können: Lasst uns eine gründliche Diskussion darüber führen, wie ein solches Projekt aufgebaut sein müsste. Ersetzt Orkut!

Doodle

Nach diesem Ausflug in den Datenschutz nun zu den nahe liegenden Problemen auf der eigenen Festplatte. Ein ganz gewöhnliches Alltagsproblem: Eine Datei liegt irgendwo, die Frage ist nur, wo und unter welchem Namen. Bekannte Kommandozeilenwerkzeuge wie »slocate« helfen nur bedingt, da sie nur nach Pfad- und Dateinamen suchen. Für die Fahndung nach Files mit einem bestimmten Inhalt muss ein anderes Werk-

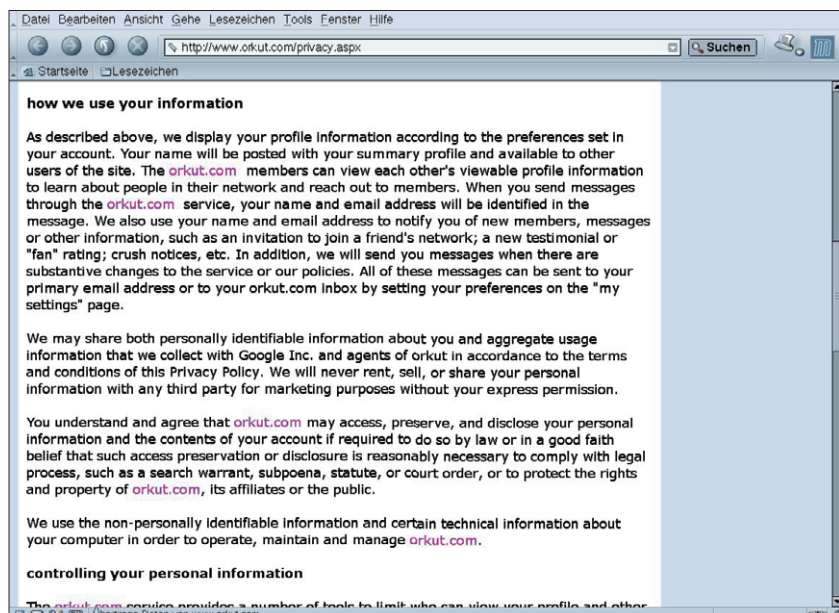


Abbildung 2: Im Privacy Statement der Orkut-Seite finden sich einige Passagen, die bei Datenschutzbeauftragten und mündigen Bürgern sicherlich auf heftige Ablehnung stoßen.

ziellen Projekten zugute: Zusätzlich zu allen anderen Funktionen gibt Pylint Statistiken aus, die dem Projektmanager dabei helfen, den Überblick über den Stand der Dinge zu bewahren.

Um die Entwicklung eines Projekts im Laufe der Zeit zu dokumentieren, legt Pylint die Auswertungsergebnisse in einer Datenbank ab. Beim nächsten Aufruf vergleicht das Programm die Ergebnisse zwischen den beiden Aufrufen. Auf diese Weise lässt sich leicht innerhalb einer Arbeitsgruppe feststellen, an welchen Teilen die Teilnehmer in letzter Zeit gearbeitet haben.

Eigenbedarf

Die Arbeit an Pylint selbst begann im April 2003; eine erste nutzbare Version lag bereits im Juni 2003 vor. Im Anschluss daran arbeitete vor allem Sylvain Thenault, der Pylint-Kernentwickler, an dem Projekt. Nach mehreren größeren Schnitten im Oktober 2003 und Februar 2004 erschien im Mai 2004 mit der Version 0.4 die erste als stabil eingestufte Fassung.

Grund für die Entwicklung war ganz klassisch der Eigenbedarf: So existierte zwar bereits ein ähnliches Projekt mit dem Namen Pychecker [9] zur automatischen Fehlersuche in Python-Projekten, aber die Pylint-Entwickler wollten etwas mehr. Speziell bei den Statistiken schien der Pychecker-Code damals nicht gut erweiterbar zu sein.

Andere Tests als in Pychecker

Pychecker führt zwar einige Tests durch, die bisher in Pylint nicht enthalten sind, jedoch liegt Pylint laut Alexandre Fayolle gerade im Bereich der Gruppenarbeit vorn: Mit dem Tool werden seiner Ansicht nach Standards stärker beachtet, nützlichere Statistiken gesammelt und die Entwicklung im Laufe der Zeit wird besser dokumentiert. Außerdem vergibt die Software eine Gesamtnote, die sich aus allen Fehlern und Warnungen im Programmcode errechnet.

Diese Gesamtbewertung blieb nicht folgenlos: Bei Logilab führte sie zunächst zu einem unerwarteten Produktivitätseinbruch, da alle Entwickler sich plötz-

lich mit vollem Enthusiasmus in ihre Lieblingsprojekte stürzten, um deren Bewertung zu verbessern.

Eine weitere Logilab-Entwicklung ist das Programm Apycot [10], der Automatic Pythonic Code Tester, auf Deutsch also ein automatischer pythonesker Codeprüfer. Er läuft bei Logilab täglich über alle Projekte und gewichtet deren Pylint-Gesamtergebnis mit dem Anteil des geprüften Codes.

Inkompatibel zu Psycho

Das Resultat ist ein erbitterter Kampf zwischen allen Projekten um den ersten Platz, bei dem zurzeit Pylint selbst die Nase vorn hat. Da Sylvain Hauptentwickler sowohl von Pylint als auch von Apycot ist, schießen – mehr im Scherz – die Verdächtigungen ins Kraut. Manipulationen wurden bisher allerdings noch nicht entdeckt.

Einige Probleme warten noch auf eine Lösung: Optimierte ein Entwickler beispielsweise das zu testende Python-Programm mit Psycho [11], ist eine Analyse nicht möglich. Psycho ist eine Art JIT-Compiler (Just in Time) für die Skriptsprache, der die Geschwindigkeit der ausgeführten Programme zum Teil erheblich steigert. Da Pylint keine Psycho-optimierten Klassen analysieren kann, behelfen sich die Entwickler zurzeit mit einem Workaround: Bei der Analyse setzen sie eine Umgebungsvariable, die vom Programm geprüft wird und so das Laden von Psycho verhindert.

Test in freier Wildbahn

Ansonsten funktioniert Pylint zuverlässig und beinahe fehlerfrei mit den bei Logilabs vorhandenen Programmen. Für die weitere Entwicklung ist jetzt der Test in freier Wildbahn wichtig, da die Entwickler weltweit unterschiedliche Python-Idiome verwenden. Nur so sind die verbleibenden Fehler zu finden.

Auf der Todo-Liste befinden sich noch einige Punkte: So sollen beispielsweise Graphen erzeugt werden, um zirkuläre Importe und Abhängigkeiten zu entdecken. Auch die grafische Benutzeroberfläche ist bisher recht sparsam gehalten. Darüber hinaus heißen die Entwickler jede Form externer Beteiligung sowie Er-

weiterungswünsche – idealerweise mit Patch – und Beiträge zur Dokumentation willkommen.

Selbstverständlich ist das Pylint-Programm vollständig in Python geschrieben und wird wie Apycot unter der GNU General Public License (GPL) als freie Software herausgegeben. Lizenzprobleme sollten einem Test also nicht im Wege stehen und die Gemeinde darf sich auf eine Menge sauber implementierten Python-Code freuen.

Damit genug für diesen Monat. Ein paar Hinweise und neue Perspektiven waren hoffentlich wieder für viele Leser dabei. Fragen, Anregungen, Kommentare, Meinungen, interessante Projekte und sonstige E-Mails bitte wieder an die übliche Adresse [12]. (agr) ■

Infos

- [1] Orkut: [<http://www.orkut.com>]
- [2] Google: [<http://www.google.com>]
- [3] GMail: [<http://gmail.google.com>]
- [4] Free-Network-Homepage: [<http://freenet.sourceforge.net/>]
- [5] Doodle: [<http://ovmj.org/doodle/>]
- [6] Libextractor-Homepage: [<http://ovmj.org/libextractor/>]
- [7] Python: [<http://www.python.org>]
- [8] Pylint-Projekt: [<http://www.logilab.org/projects/pylint>]
- [9] Pychecker: [<http://pychecker.sf.net>]
- [10] Apycot-Projekt: [<http://www.logilab.org/projects/apycot/>]
- [11] Psycho-Homepage: [<http://psycho.sf.net>]
- [12] Kritik und Anregungen: [column@brave-gnu-world.org]
- [13] GNU-Projekt: [<http://www.gnu.org>]
- [14] Homepage der Brave GNU World: [<http://brave-gnu-world.org>]
- [15] „We run GNU“-Initiative: [<http://www.gnu.org/brave-gnu-world/rungnu/rungnu.de.html>]

Der Autor

Dipl.-Phys. Georg C. F. Greve beschäftigt sich seit etlichen Jahren mit freier Software und kam früh zu GNU/Linux. Nach Mitarbeit im GNU-Projekt und seiner Aktivität als dessen europä-



ischer Sprecher hat er die Free Software Foundation Europe initiiert, deren Präsident er ist. Mehr Informationen finden sich unter: [<http://www.gnuhh.org>]