

Yasni (Interview I): Die Personensuchmaschine

(Darmstadt, 17. Februar 2010)

Das Ausmaß, die Vielfalt und die Anzahl der eingebundenen Quellen von Yasni werden erst durch Recherchen sichtbar. In diesem Interview möchte ich deswegen meinen (Fach-)Lesern einen kleinen Einblick in die Personensuche mit Hinblick auf Indizes, Quellen und Trefferrelevanz geben. Weitere Punkte sind die Usability und speziellen Usecases bei Yasni.

SQ: Herr Rühl (SR), nach welchen Kriterien werden die Quellen für Yasni ausgewählt?

SR: Die Quellen werden nach Bedeutung für das jeweilige Land ausgewählt. Yasni ist außer für Deutschland, Österreich und der Schweiz auch mit einer Version für UK und einer internationalen Version verfügbar.

Zudem müssen sie grundsätzlich öffentlich verfügbare Informationen bereithalten. Für bestimmte Bereiche wie z. B. Bilder und Videos fragt Yasni selbst Metasuchmaschinen ab, so dass eine sehr breite Abdeckung sichergestellt wird.

SQ: Yasni zeigt nicht nur die gefundenen Dokumente (Texte, Adressen, Bilder etc.) zu einer Person an, sondern man kann sie auch gezielt zu dem eigenen Profil hinzufügen oder entfernen – bei Yasni! Nicht zu verwechseln mit „Löschen aus dem Google-Index“!

SQ: Welche Rolle spielt bei der Personensuche auch die Aktualität der Indexe und ab wann wird ein Ergebnis nicht mehr angezeigt?

SR: Yasni hat zwar wie jede Suchmaschine auch einen Zwischenspeicher, um die abgefragten Websites nicht übermäßig zu belasten, allerdings werden die Informationen bei einer Suche in Echtzeit aktualisiert, sofern sie ein bestimmtes Alter erreicht haben.

Je nach Art der Information kann dieses „Haltbarkeitsalter“ zwischen 1 Stunde und 2 Wochen liegen. Das Prinzip ist ähnlich dem von Google, das ja auch Newsquellen häufiger aktualisiert als z.B. dauerhaft gültige biografische Informationen.

SQ: Als Diplom Informationswirtin habe ich nicht nur Indizes, Auswahl und Erschließung der Quellen, sondern auch Suchwortzerlegung untersucht. Das heißt die Suche wurde ordentlich gequält, z. B. durch besondere Namen, die jede Suche erstmal verdauen muss, z. B.

- Namen mit Zusätzen, wie den meiner Kollegin „Eva Rausch von Traubenberg“ (link-baronesse.de),
- Doppelnamen die sowohl Vornamen, als auch Nachnamen sein können „Hans Joachim Gabriel“,
- und Fehlertoleranz.

SQ: Wie „pimpen“ Sie die unterschiedliche Ergebnisqualität der eingebundenen Quellen und gehen z. B. mit den Besonderheiten von Namen um?

SR: Yasni nutzt einen selbst entwickelten Algorithmus für die Relevanzberechnung, der sich neben der Art der Quelle auch auf die Information selbst und die Aktualität der Information bezieht. Ziel ist, dem Nutzer abhängig vom gesuchten Namen möglichst die von ihm erwartete Reihenfolge anzuzeigen.

Quellen und Bereiche, die keine Ergebnisse liefern werden – anders als bei anderen Suchmaschinen – komplett ausgeblendet, damit der Nutzer keine „irrelevante Information“ erhält.

SQ: Das heißt, sie verwenden keine Standard-Suchmaschinentechnologie, wie z. B. FASTSearch

SR: Ja, das ist richtig! Wir haben eine Eigenentwicklung am Start.

SQ: Erfahrungsgemäß ist eine manuelle Nachbearbeitung bei so vielen aggregierten Quellen wie bei Yasni an der Tagesordnung. Herr Schütz (FS), welche Kriterien bzw. Regeln hierfür verwenden Sie, um die Qualitätssicherung (QS) einfacher zu gestalten?

FS: Alle Mitarbeiter testen neben dem festen Testteam von Zeit zu Zeit mit, Feedback unserer User wird ausgewertet und wir hören Blogautoren, Freunden und Geschäftspartnern genau zu.

Yasni soll für jeden Nutzer „gute Ergebnisse“ liefern – das funktioniert nicht mit einem einfachen Regelset, da abhängig vom Suchenden, dem gesuchten Namen und der Suchmotivation die Qualität immer wieder anders beurteilt wird.

Beispielsweise wird eine mit Regeln auf Web2.0-User getrimmte Suche niemals gute, im Sinne von relevante, Ergebnisse für die Masse der Nutzer liefern.

Kurzantwort: die QS ist NICHT einfach gestaltet.

SQ: Gibt es auch Quellen, die aufgrund der geringen Relevanz ausgeschlossen wurden?

FS: Ja, überwiegend die Websites, die rein zur Suchmaschinenoptimierung und ohne Mehrwert für Nutzer gestaltet wurden. Diese schließen wir aus. Zudem zeigen wir in Verzeichnissen eindeutig mehrfach vorkommende Einträge (Dubletten) nicht mehrfach in der Ergebnisseite an, sondern entscheiden uns hier in der Tat für den qualitativ besten Anbieter.

SQ: Bei welchen Diensten wird wie nachgearbeitet, um eine gute Trefferqualität zu erreichen?

SR: Alle Quellen werden mit sorgfältig ausgewähltem Abfragemuster und in der Regel mit einer automatisierten Nachfilterung verarbeitet. Diese Funktionen gehören im weitesten Sinn auch zur Relevanz-Berechnung.

SQ: Das heißt, sie führen regelmäßig Relevanztests durch. Wie sieht die Methodik dazu aus?

SR: Regelmäßige und wiederholbare Tests werden automatisiert, sowie manuell mehrfach täglich nach verschiedenen methodischen Ansätzen durchgeführt. Wir optimieren dadurch stetig unsere Trefferqualität und Abfragealgorithmen.

Wir haben Mitarbeiter, die sich ausschließlich nur darum kümmern! Im Klartext: Suchanfragen, Fragestellungen und Usecases nachstellen und die Ergebnisse nach Relevanz überprüfen. Dadurch lässt sich eine hohe Qualität und Relevanz der Ergebnisse für unsere Personensuche nachhaltig gewährleisten.

SQ: Gibt es bei Yasni eine Aufteilung im Sinne von Kategorisierung der gestellten Suchen, z. B. in Personen, Städte, Unternehmen, Telefonnummern?

FS: Ja! Yasni unterscheidet sich gerade darin von herkömmlichen Personensuchmaschinen, DASS gezielt nach Kategorien gesucht werden kann, oder auch frei nach Namensbestandteilen, z. B. in Verbindung mit einer Stadt oder Firma.

Die Suche nach Namen oder namentlich bekannten Personen überwiegt zurzeit noch, aber rasant wächst der Anteil der Suchen nach Personen zu z. B.:

- Orten (Wer wohnt in ...?)
- Begriffen (DSDS = Wer sind die Teilnehmer?)
- Ingenieur (BMW = Wer arbeitet als Ingenieur bei BMW?)
- Kombinationen (Herr Müller bei Google)

SQ: Welche Kategorien liegen vor?

FS: Man kann man ganz beliebige Begriffe miteinander kombinieren und danach suchen – Google hat ja auch keine festen Kategorien.

Wichtige Typen wie Namensbestandteil, Firmenname, Ort oder Beruf werden automatisch erkannt und diese Information für ein noch besseres Ergebnis genutzt.

Probieren Sie doch mal die Suche „Frau Quirmbach von Deutsche Telekom“ bei Yasni und Sie sehen unter „Meinten Sie?“ dass die Bestandteile der Suche korrekt erkannt werden.

SQ: Habe ich gemacht und wird auch richtig angegeben! Die Rechtschreibkorrektur hat mir allerdings etwas gefehlt. Es gibt zwar die oben beschriebene Möglichkeit noch mal die Namensbestandteile in feste Kategorien einzuordnen, jedoch kein „Meinten Sie ...?“ im eigentlichen Sinne, wie das z. B: Google macht.

Ist zugegeben auch etwas komplizierter, da man zuerst einmal einen Namens-, Unternehmens- und Gewerkelisten zum Abgleichen haben muss. Durchaus auch eine Kostenfrage, so ein Feature anzubieten. Besteht nämlich keine Möglichkeit die Listen selbst zu generieren, müssen sie extern eingekauft werden.

SQ: Haben Sie dennoch ein solches für Suchende sehr hilfreiches Feature vorgesehen?

SR: Natürlich werden wir auch dieses Feature anbieten! Wir arbeiten ständig an weiteren Verbesserungen und Erweiterungen der Personensuche und stehen hier sicher noch ganz am Anfang einer komplett neuen Kategorie von Suchmaschinen.

SQ: Wird je nach Suchanfrage eine unterschiedlich aufgebaute SERP (Search Engine Result Page) ausgeliefert?

SR: Ja. Es gibt grundsätzlich die Ergebnisarten:

- „gefundene Personen und Namen zu einer Personensuche“ und
- „Personen und Webergebnisse zu einer Namensrecherche“.

Beide SERPs basieren aufeinander – wenn Ihnen bei den Ergebnissen zu Ihrem Namen oder in Ihrem Profil Begriffe in der Begriffsbox rechts angezeigt werden, werden Sie zu diesen Begriffen in der Personensuche auch gefunden natürlich auch zu vielen weiteren Begriffen.

Es gibt also 2 Ausgangs-Entitäten bzw. 2 SERPS:

- nach Name
- nach Begriff

Beide SERPS sind jeweils unendlich granular und beliebig erweiterbar.

SQ: Verwenden Sie eine Blacklist, um Inhalte zu filtern und bewusst nicht mit in der SERP auszugeben, z. B. bei pornografischen Anfragen?

FS: Es gibt eine Kontrolle der von Nutzern eingestellten Inhalte auf jugendgefährdende Bilder und Begriffe und bestimmte eindeutig strafbare Suchkombinationen sind gesperrt.

Wie jede Suchmaschine wird natürlich auch Yasni für alle Bedürfnisse, auch Dating und direkte sexuelle Anregungen, genutzt und liefert hier oft auch gute Ergebnisse.

Die Grenzen zwischen prominentem Model, „seriösem Dating“ und Pornografie sind allerdings fließend und wir lehnen auch hier jede Form der Zensur ab. Zu weit überwiegendem Anteil wird Yasni für „normale Namen und Begriffe“ und völlig seriöse Zwecke genutzt.

SQ: Welche Erfahrungen zum Nutzungsverhalten in Yasni haben Sie bisher gemacht und welche Likes/Dislikes haben sich daraus ergeben?

SR: Es existieren im Grunde so viele Suchmotivationen unseres Produktes wie Nutzer!

Sehr gern wird nach Namensbestandteilen gesucht in Verbindung mit Firmen und Orten (jeder kennt sicher die Situation in der man nicht mehr den vollständigen Namen kennt) und bei aktuellen Ereignissen natürlich nach den betreffenden Personen.

Zunehmend wird aus professionellen Gründen gesucht mit Kombinationen aus Beruf/Ort oder Dienstleistungsangebot/Ort. Es gibt in der SERP die Möglichkeit noch mal die Suche zu verfeinern, indem explizit jeder Namensbestandteil einer Kategorie zugeordnet werden kann, wie Name oder Ort.

Sehr sensibel sind die Nutzer allerdings bei der direkten Anzeige ihrer Kontaktdaten wie Email oder Telefonnummer! Yasni verzichtet daher auf die technisch einfach machbare Separierung und explizite Anzeige dieser Daten, verweist aber stattdessen auf entsprechende Telefonverzeichnisse.

Generell verstehen wir uns nicht als Anbieter der Information, sondern als eine Art „Wegweiser“.

SQ: Vielen Dank für das Interview!