**بسمه تعالی**

**مستندات پروژه پایانی کارشناسی** با موضوع **طراحی و پیاده سازی موتور جستجوگر وب**

استاد راهنما: دکتر مهدی یعقوبی

دانشجو: مسلم آخوندی

دانشکده فنی مهندسی گرگان – دانشگاه گلستان

رشته مهندسی کامپیوتر – گرایش مهندسی نرم افزار

**فهرست مطالب:**

|  |  |
| --- | --- |
| آشنایی با موتورهای جستجوگر |  |
| توضیحات پروژه |  |
| بخش خزنده وب (WebCrawler) |  |
| بخش ایندکسر (Indexer) |  |
| بخش رتبه بندی (Ranker) |  |
| بخش جستجو (Search) |  |

**آشنایی با موتورهای جستجوگر وب**

موتور جستُجوی وب به انگلیسی Web search engine  ابزاریست که به منظور جُستجو در وب برای بدست‌آوردن اطلاعات درخواست شده، به کار می‌رود. نتایج یافته شده به‌طور معمول در صفحه‌ای با عنوان صفحهٔ نتایج جستجو(Search engine results page) فهرست می‌شوند.

با استفاده از کلمهٔ کلیدی(keyword) که در واقع توضیحی است کوتاه دربارهٔ آنچه لازم است در اینترنت پیدا شود، کلمه کلیدی باید تا آنجا که ممکن است کوتاه، جزیی و دقیق باشد. به غیر از وارد کردن کلمهٔ مستقیم، می‌توان با استفاده از عملگرهایی عمل جستجو را دقیق تر و منظم تر انجام داد.

**چگونگی عملکرد موتورهای جستجوی وب**

موتورهای جستجوی وب برای آنکه بتوانند به درخواست‌ها و جستجوهای کاربران پاسخ مناسبی بدهند باید دو کار مهم را انجام دهند:

* **خزیدن در وب‌سایت‌ها و ایندکس کردن صفحات**

موتورهای جستجو با استفاده از خزنده‌های خود می‌توانند تمامی صفحات و فایل‌های موجود در وب را ایندکس کنند. این خزنده‌ها با ورود به هر صفحه به ‌دنبال لینک‌ها هستند تا با دنبال کردن آن‌ها وارد صفحات جدید شوند. با این روش، موتورهای جستجو می‌توانند تمامی صفحات موجود در وب را ایندکس کنند.

* **رتبه‌بندی وب‌سایت‌ها و نمایش بهترین نتایج**

اصلی‌ترین وظیفهٔ موتورهای جستجوی وب، ارائهٔ بهترین و مرتبط‌ترین نتایج به کاربران است. موتورهای جستجو برای آنکه بتوانند بهترین نتایج را هنگام جستجوی یک عبارت خاص به کاربران نمایش دهند، قوانین و استانداردهایی برای وب‌سایت‌ها تعریف کرده‌اند تا بتوانند علاوه بر دسترسی ساده‌تر به صفحات وب‌سایت‌ها، محتوای موجود در صفحات را بهتر درک کنند. با این کار، موتورهای جستجو می‌توانند وب‌سایت‌هایی که دارای بیشترین ارتباط معنایی با عبارت جستجو شده توسط کاربر هستند در رتبه‌های بالاتری در نتایج جستجو قرار دهند.

موتورهای جستجو برای ایجاد نتیجه یک جستجو، تعدادی فعالیت انجام می‌دهند:

خزش یا Crawling فرایند واکشی تمام صفحات وب لینک شده به یک وب سایت؛ که با استفاده از نرم‌افزارهایی به نام خزشگر (Crawler)انجام می‌شود. خزشگر گوگل Google bot نام دارد.

اندیس گذاری یا Indexing فرایند ایجاد اندیس برای تمام صفحات واکشی شده و ذخیره و نگهداری این اندیس‌ها در یک پایگاه داده بسیار بزرگ که در آینده استفاده شوند. به عبارتی، فرایند اندیس گذاری از یک طرف شناسایی کلمات و عباراتی است که صفحه را به خوبی توصیف می‌کند و از طرف دیگر انتساب کلمات کلیدی یافت شده به صفحه می‌باشد.

پردازش یا Processing وقتی درخواست جستجو به موتور جستجو می‌رسد، موتور جستجو آن را پردازش می‌کند. در واقع موتور جستجو رشته جستجو را با صفحات اندیس شده در پایگاه داده مقایسه می‌کند.

محاسبه ارتباط یا Relevancy از آنجایی که ممکن است بیشتر از یک صفحه حاوی رشته جستجو باشند، بنابراین موتور جستجو شروع به محاسبه ارتباط هر صفحه با رشته جستجو می‌کند. برگرداندن نتیجه: آخرین گامی که موتور جستجو انجام می‌دهد، بازیابی بهترین نتیجه منطبق می‌باشد. به عبارت ساده، این فعالیت چیزی نیست جز نمایش اطلاعات در مرورگر.

**توضیحات پروژه**

موضوع پروژه: طراحی و پیاده سازی موتور جستجوگر وب

لینک گیت هاب: https://github.com/ariantron/WebSearchEngine

**بخش ها:**

|  |  |  |  |
| --- | --- | --- | --- |
| **ردیف** | **عنوان** | **زبان برنامه نویسی** | **نرم افزار پایگاه داده** |
| ۱ | خزنده وب (WebCrawler) | Java | MySQL |
| ۲ | ایندکسر (Indexer) | Java | MySQL |
| ۳ | رتبه بندی (Ranking) | Java | MySQL |
| ۴ | جستجو (Search) | PHP (فریمورک لاراول) | MySQL |