

پیاده سازی موتور جستجو



بازیابی اطلاعات و جستجوی وب

استاد: دکتر محمود نشاطی

امیر شمس علیئی - ۹۲۲۱۳۱۲۷

صالح جعفریزاده - ۹۲۲۱۳۰۶۱

مراحل انجام پروژه

انتخاب
سایت‌های
مناسب برای
شروع خزش

راه اندازی
Crawler

- استفاده از Crawler4j

بررسی قالب
سایت‌های
انتخاب شده

- انتخاب regex مناسب
برای url ها
- بررسی ساختار درج آگهی

ایندکس کردن
فیلدهای
مربوط به آگهی

- استفاده از lucene

پیاده سازی
متودهای
جستجو

- جستجو بر اساس
فیلدهای مختلف

چالش‌های کار

- ساختار نامناسب بعضی از سایت‌ها
 - مشخص نبودن اسم تگ مربوط به متن استخدام
 - مشخص نبودن مکان استخدام
 - مشخص نبودن زمان درج آگهی
- تشخیص سایت‌های مناسب و قابل اعتماد برای خزش
 - مسدود شدن بعضی از سایت‌ها بعد از خزش اولیه و قبل از خزش نهایی
 - به روز نبودن بعضی از سایت‌ها

سایت های خزش شده

www.hamikar.com	• 999	dehvand.ir/	• 81
mazandkar.ir	• 6	karsaz-esf.ir/component/jbjobs/guest/joblist?Itemid=102	• 41
karfa.ir/	• 30	www.cityad.ir/	• 330
www.irankar.biz	• 0	www.karkon.org/	• 634
job.pnuna.com	• 508	karjob.com/jobs	• 47
estekhdamjoo.ir	• 26	www.nofa.ir/	• 100
www.estekhdami.org/	• 391	www.e-estekhdam.com/	• 942
www.estekhdami.com/	• 35	karbank.ir/	• 11
www.estekhdamiran.com/	• 56	www.jobcity.ir/	• 41
karan.ir/request/index	• 1	ekaar.ir/	• 34

پیشنهاد هایی برای راه اندازی موتور جستجو

- ارزش گذاری برای سایتها جهت خزش بر اساس بازخورد کاربران
 - بازگشت به صفحه جستجو بعد از انتخاب یک نتیجه
 - جستجوی عبارت مشابه در بازه زمانی کوتاه بعد از جستجوی اول
- در نظر گرفتن امکانی برای معرفی سایتها به خزشگر توسط صاحبان سایت و کاربران
- تهیه استاندارد ساختار سایت و قالب صفحه برای خزش بهتر و آسان تر
 - سایتها برای اینکه رتبه بهتری در جستجو داشته باشند از این امکان استفاده کنند

“
چو غنچه گرچه فروبستگی است، کار جهان
تو همچو باد بهاری گره گش می باش
”

لسان الغیب حافظ شیرازی

با تشکر از همراهیتان!

