

(۱) بازیابی اطلاعات چیست؟

سوال: بازیابی اطلاعات را تعریف کنید.

جواب: بازیابی اطلاعات یعنی یافتن و مرتب‌سازی اسناد/متن‌های مرتبط با نیاز یک کاربر.

(۲) تفاوت جستجو و بازیابی اطلاعات؟

سوال: جستجو و بازیابی اطلاعات چه تفاوتی دارند؟

جواب: جستجو معمولاً فقط نتایج را پیدا می‌کند، اما بازیابی اطلاعات علاوه بر پیدا کردن، مرتبط‌بودن را هم در نظر می‌گیرد و معمولاً نتایج را رتبه‌بندی می‌کند.

(۳) نیاز کاربر چه نقشی در بازیابی دارد؟

سوال: نیاز کاربر چگونه بر خروجی سیستم اثر می‌گذارد؟

جواب: نیاز کاربر با یک پرسش/عبارت نشان داده می‌شود و سیستم بر اساس آن، اسناد مرتبط را پیدا و رتبه‌بندی می‌کند.

(۴) پرسش (عبارت جستجو) چیست؟

سوال: پرسش را تعریف کنید.

جواب: پرسش یعنی مجموعه کلمات یا عباراتی که کاربر برای پیدا کردن اسناد مرتبط می‌نویسد.

(۵) معیار «دقت» را تعریف کنید.

سوال: دقت چیست؟

جواب:

دقت = تعداد اسناد مرتبط بازیابی شده / تعداد کل اسناد بازیابی شده
تبدیل بازیابی شده

(۶) معیار «بازیابی» را تعریف کنید.

سوال: بازیابی چیست؟

جواب:

بازیابی = تعداد اسناد مرتبط بازیابی شده / تعداد کل اسناد مرتبط
بازیابی شده

(۷) مدل بولی چه ایده‌ای دارد؟

سوال: مدل بولی در بازیابی اطلاعات چه کاری انجام می‌دهد؟

جواب: در مدل بولی هر سند یا شرطها را برآورده می‌کند یا برآورده نمی‌کند (در قالب بله/خیر) و از عملگرهایی مثل «و» و «یا» و «نه» استفاده می‌شود.

(۸) بسامد واژه در یک سند چیست؟

سوال: بسامد واژه یعنی چه؟

جواب: بسامد واژه یعنی تعداد دفعاتی که یک کلمه در یک سند تکرار شده است.

(۹) توکن‌سازی چیست؟

سوال: توکن‌سازی را توضیح دهید.

جواب: توکن‌سازی یعنی شکستن متن به واحدهای کوچک‌تر (مثل کلمه‌ها) تا بتوان راحت‌تر پردازش کرد.

(۱۰) چرا واژه‌های بی‌اهمیت حذف می‌شوند؟

سوال: حذف کلمات بی‌اهمیت (مثل «و»، «در»، «از») چه فایده‌ای دارد؟

جواب: باعث کاهش نویز می‌شود و کمک می‌کند سیستم روی کلمات مهم‌تر تمرکز کند و کیفیت بازیابی بهتر شود