



سوال ۱: فشرده سازی (۳۰ نمره)

برای سوالات زیر پاسخ کوتاه بنویسید.

- فرض کنید posting list یک کلمه در یک داک، به شکل زیر باشد:

[6, 29, 170, 238]

با استفاده از روش فشرده سازی variable byte این posting list را فشرده کنید.

- فرض کنید posting list یک کلمه در یک داک، به شکل زیر و با روش gamma code فشرده شده است:

1100111100011111000001111110000010

با توضیح مراحل میانی، posting list اولیه را به دست آورید.

سوال ۲: رتبه بندی سیستم های بازیابی (۲۵ نمره)

فرض کنید یک سیستم بازیابی رتبه بندی زیر را از چپ به راست ارائه داده است:

R, N, R, N, N, R, N, N, R, R

که R نشان دهنده مرتبط بودن است و N نشان دهنده غیر مرتبط بودن. نمودار P-R curve را برای این سیستم بکشید و همچنین mean average precision را محاسبه کنید.

سوال ۳: امتیاز دهی در فضای برداری (۴۵ نمره)

فرض کنید یک سیستم بازیابی اطلاعات با مدل Inc.ltn دارید. همچنین دو document و یک query به شکل زیر به شما داده شده است. با توجه به معیار گفته شده، امتیاز query را برای هر یک از documentها محاسبه کنید. فرض کنید کل corpus تنها شامل این دو document می شود. همچنین مراحل میانی را از جمله به دست آوردن ... tf, idf, بنویسید.

Doc1: hi good course laugh all time

Doc2: bye take life easy it is an easy course

Query: laugh to life easy course