

(ج) (۷ نمره) S زیرفضایی از V میباشد همبطنور مجموعه‌ای از اعضای V را به صورت $B = \{b_1, \dots, b_k\}$ را در نظر بگیرید که $\forall v \in V v \in \text{Span}(B)$.
درستی عبارت زیر را بسنجید.

$$S = \bigoplus_{i=1}^k (S \cap \text{Span}(b_i))$$

(د) (۷ نمره) برای زیرفضاهای W_1, \dots, W_k در V به طوری که $W_i \neq V$ نشان دهید که عضوی در V وجود دارد که در هیچ یک از W_i ها نمیباشد.

پرسش ۵ (۲۸ نمره) برای تمامی مقادیر k مشخص کنید که چه زمانی دستگاه معادلات زیر جواب ندارد، جواب یکتا دارد و یا بینهایت جواب دارد.

$$\begin{aligned} x + 5y - 3z &= 2 \\ -2x - 7y + 3z &= -5 \\ -x - 5y + (k^2 - 6)z &= k + 1 \end{aligned}$$

سوالات عملی (۳۲ نمره)

تاریخ تحویل: ۸ اسفند ۱۴۰۱

پرسش ۱ (۳۲ نمره) همان‌طور که میدانید، کامپیوتر موجود بسیار نفهمی است و تمام تلاش خود را برای نفهم بودن می‌کند. یکی از مصادیق نفهم بودن کامپیوتر در نفهمیدن متون و شباهت میان آنها است. این است که برای فهماندن متن به کامپیوتر، چاره‌ای به جز تولید یک بازنمایی برداری از آن نداریم. یک بازنمایی برداری معروف در کامپیوتر از متون، کد ASCII و یا UTF برای ذخیره‌سازی کاراکترها و استفاده از دنباله‌ای از آنها برای ذخیره‌سازی یک متن است. اما مشکلی که این بازنمایی دارد این است که هیچ ارتباطی به بعد معنایی کلمات و جمله نخواهد داشت.

می‌خواهیم یک بازنمایی جدید برای جملات معرفی کنیم. فرض کنید که مجموعه‌ای از m جمله داریم که در مجموع از n کلمه متمایز تشکیل شده‌اند. حال از روی کلمات و جملات می‌خواهیم دو ماتریس زیر را تعریف کنیم.

(آ) ماتریس بسامد کلمات: هر چه یک کلمه در یک جمله پرتکرار تر باشد، احتمالاً نقش پررنگ‌تری در معنی آن جمله خواهد داشت. ماتریس بسامد کلمات که آن را F می‌نامیم یک ماتریس $m \times n$ است که درایه j, i آن برابر بسامد کلمه j در جمله i است. به عنوان مثال در جمله

I like eating desert in desert.

۶ کلمه داریم و کلمه desert در آن دو بار تکرار شده است. بنابراین بسامد این کلمه در این جمله برابر $\frac{2}{6}$ است.

(ب) ماتریس متداول بودن کلمات: هر چه یک کلمه در جملات بیشتری واقع شده باشد، احتمالاً بار معنایی کمتری دارد. به عنوان مثال حروف اضافی و ربط در اکثر قریب به اتفاق جملات واقع شده‌اند. ماتریس (بردار) تداول کلمات یک ماتریس $1 \times n$ است که درایه i ام آن برابر لگاریتم نسبت تعداد کل جملات به تعداد جملاتی است که شامل کلمه i ام باشند. به صورت رسمی اگر D مجموعه کل جملات باشد و D_i زیر مجموعه از جملات باشد که شامل کلمه i ام باشند، درایه i ام ماتریس تداول کلمات که با I آن را نمایش می‌دهیم، برابر است با

$$I[i] = \ln \frac{|D|}{|D_i|}$$

حال بازنمایی جمله i ام را به صورت ضرب درایه به درایه ردیف i ام ماتریس بسامد در بردار تداول در نظر می‌گیریم.

در این مساله یک مجموعه به نام S داریم که شامل تعدادی جمله است. در مرحله اول باید بازنمایی این جملات محاسبه شود. سپس یک جمله query داریم که باید بازنمایی آن مطابق بازنمایی محاسبه شده برای جملات بالا محاسبه شود. در نهایت باید اندیس شبیه‌ترین جمله‌ی S به query برگردانده شود. این شباهت از طریق محاسبه زاویه بدست می‌آید.

پیش از انجام هر یک از این موارد، باید علائم نگارشی متن را حذف کنید. همچنین باید از کوچک یا بزرگ بودن حروف صرف نظر شود.

ورودی

در خط اول ورودی عدد صحیح n وارد می‌شود که برابر $|S|$ است. در n خط بعدی در هر خط یک جمله از مجموعه S وارد می‌شود. در نهایت در خط بعدی جمله query وارد می‌شود.

خروجی

در خروجی باید اندیس جمله‌ای که بیشترین شباهت را به query دارد چاپ شود.

ورودی نمونه

```
3
This is the first document.
This document is the second document.
And that is the third one.
Is this the first document?
```

خروجی نمونه

```
1
```