

# پنج‌شنبه‌ی پنج‌م

چند نکته در مورد رقابت امروز: یکم، در رقابت امروز سریع‌ترین جواب ارسال شده، بهترین جواب خواهد بود. اگر اختلاف زمان اجرای دو جواب کمتر از دو دهم ثانیه باشد، جوابی که زودتر فرستاده شده است بهترین جواب در نظر گرفته می‌شود. دوم، برنامه‌ی داور از این پس زبان جاوا را پشتیبانی می‌کند و جواب‌ها می‌توانند در زبان جاوا نیز نوشته شوند. سوم، بهتر است نام کاربری، قسمتی از نام و نام خانوادگی شرکت‌کننده را در بر داشته باشد تا بتوان با توجه به آن برنده را تشخیص داد. در پایان، اعضای تیم فنی رقابت پنج‌شنبه‌های سخت، برای شما بهترین لحظات و پیروزی را در این رقابت آرزو می‌کنند.

عنوان مسئله	یادداشت‌های خیس
شناسه‌ی مسئله	ct05
سختی مسئله	۵ از ۹
زمان شروع	ساعت ۱۶ ۱۳۹۵/۷/۲۹
زمان پایان	ساعت ۱۶ ۱۳۹۵/۸/۴

## بیان مسئله

## یادداشت‌های خیس

مسئولیت جمع‌بندی و یادداشت مطالب مطرح شده در جلسه‌های تیم فنی صد و یک نفره‌ی پنج‌شنبه‌ها، مدتی است که بر دوش آقای دال (حرف اول نام یکی از اعضای محترم این تیم) نهاده شده است. چون نرخ نامه‌های دریافت شده توسط اعضای محترم تیم پنج‌شنبه‌ها افزایش یافته است، اهمیت این یادداشت‌ها و انتشار آنها نیز بیشتر شده است. از این رو، دال تمام تلاش خود را انجام می‌دهد تا این مسئولیت را به خوبی انجام دهد. متأسفانه در باران شدید هفته‌ی گذشته، برخی از یادداشت‌های دال خیس شده‌اند و نوشته‌های آن به سختی قابل خواندن هستند. دال با کمال اندوه و تأسف از این اتفاق و پس از ساعت‌ها بررسی، توانست کلمه‌های این یادداشت‌ها را استخراج کند اما نقطه‌های این کلمه‌ها به سختی قابل تشخیص هستند. پس از ناتوانی از حدس زدن مکان نقطه‌های این کلمه‌ها، دال به این نتیجه رسید که این یادداشت‌ها غیر قابل بازیافت هستند و بشریت از چه دستاوردهایی محروم شده‌اند که ناگهان نامه‌ی زیر توجه او را به خود جلب کرد.

آقای دال عزیز،

شهرت پنج‌شنبه‌ها و بستر بی نظیری که این رقابت برای حل سخت‌ترین مسئله‌های جهان فراهم می‌کند، سراسر کشور را فرا گرفته است. ما نیز پس از ناتوانی از یافتن جوابی راضی‌کننده برای یکی از چالش‌های پیش رو در تیمی که من عضو آن هستم، تصمیم گرفتیم با شما مکاتبه نماییم. برای افزایش سرعت نوشتن فارسی با استفاده از نمایشگرهای لمسی ما راهکارهای متفاوتی را بررسی کرده‌ایم. یکی از ایده‌های پیشنهاد شده، کاهش تعداد کلیدهای لازم برای نوشتن حروف فارسی است. در این ایده، برای همه‌ی کلمه‌های هم‌شکل مثل «ح» و «ج» فقط یک کلید در نظر گرفته می‌شود. پس از پایان هر کلمه، با استفاده از یک لغت‌نامه، کلمه‌ی مورد نظر از بین کلمه‌های «مشابه» انتخاب می‌شود. دو کلمه مشابه هستند اگر اختلاف آنها فقط در نقطه‌هایشان باشد. برای مثال، پس از وارد کردن کلمه‌ی «سبر» از بین کلمه‌های «سپر»، «سبز»، «شیر» و «سیر»، یکی انتخاب می‌شود. ما قصد داریم تعداد کلمه‌های مشابه‌ی هر کلمه را محاسبه کنیم تا با توجه به آن، این ایده را ارزیابی نماییم. لطفاً به ما کمک کنید.

با احترام فراوان،

سین

با خواندن این نامه، روزنه‌های امیدی در ذهن دال پدیدار شده‌اند: به ازای هر کلمه‌ای که از یادداشت‌ها استخراج شده است، اگر نزدیک‌ترین کلمه‌ی مشابه (با توجه به مکان نقطه‌ها) انتخاب گردد، احتمالاً بیشتر کلمه‌ها به درستی تصحیح می‌شوند. اما این راه در صورتی به نتیجه خواهد رسید که نزدیک‌ترین کلمه‌های مشابه برای هر یک از کلمه‌های استخراج شده محدود باشند. به دال برای پاسخ به این پرسش کمک کنید.

در این مسئله، کلمه‌های لغت‌نامه و تعدادی کلمه‌ی پرسش داده می‌شوند. به ازای هر کلمه‌ی پرسش، باید تعداد نزدیک‌ترین کلمه‌های مشابه موجود در لغت‌نامه محاسبه گردد. برای مثال اگر کلمه‌ی پرسش «سبر» باشد و چهار کلمه‌ی «سپر»، «سبز»، «شیر» و «سیر» در لغت‌نامه موجود باشند، اختلاف نقاط کلمه‌ی پرسش با این کلمه‌ها به ترتیب دو، یک، چهار و یک می‌باشد. بنابراین دو کلمه هستند که فقط در یک نقطه با کلمه‌ی پرسش تفاوت دارند. به بیان دیگر، تعداد نزدیک‌ترین کلمه‌های مشابه به «سبر» دو است.

توجه به این نکات لازم است: یکم، قرار گرفتن نقطه پایین یا بالای حروف اهمیت دارد. برای مثال اختلاف نقاط «پر» با «بر» دو نقطه است ولی اختلاف «پر» با «تر» پنج نقطه است. دوم، حروف «ی» و «ن» وقتی با حروف خانواده‌ی «ب» مشابه می‌شوند که در آخر کلمه ظاهر نشوند. برای مثال «یا» و «تا» با هم مشابه هستند ولی «کی» و «کت» نیستند (شکل دو حرف «ن» و «ی» در آخر کلمه‌ها با سایر حروف این خانواده متمایز می‌شود). همین موضوع برای حروف «ق» و «ف» نیز صدق می‌کند و این دو حرف در صورتی مشابه هستند که در پایان یک کلمه ظاهر نشوند. سوم، ورودی‌ها به صورت UTF-8 خوانده می‌شوند.

در فایل <http://strongspace.com/ali/public/ct05.txt> همه‌ی حروفی که در کلمه‌های لغت‌نامه و پرسش می‌توانند ظاهر شوند، نوشته شده‌اند. در این فایل، حروف مشابه در یک خط قرار داده شده‌اند.

## نمونه‌های ورودی و خروجی

ورودی با دو عدد شروع می‌شود. عدد اول تعداد کلمه‌های لغت‌نامه (حداکثر پنجاه هزار) و عدد دوم تعداد کلمه‌های پرسش (حداکثر پنج هزار) را نشان می‌دهند. سپس کلمه‌های لغت‌نامه ظاهر می‌شوند (هر یک در یک خط). پس از آن کلمه‌های پرسش قرار می‌گیرند (هر یک در یک خط). هر یک از این کلمه‌ها حداکثر بیست حرف دارند. خروجی باید به تعداد کلمه‌های پرسش خط داشته باشد که در هر خط، تعداد نزدیک‌ترین کلمه‌های مشابه کلمه‌ی پرسش متناظر با آن خط (از بین کلمه‌های لغت‌نامه) را نشان می‌دهد.

یک نمونه‌ی ورودی و خروجی در ادامه مشاهده می‌شود. در این نمونه، نزدیک‌ترین کلمه‌های مشابه «سبر»، «سبز» و «سیر» هستند و مشابه «شیر» خود «شیر» است. همچنین، کلمه‌ای در لغت‌نامه مشابه «تیز» نیست.

ورودی	خروجی
4 3	2
سیر	0
شیر	1
سبز	
سپر	
سبر	
تیز	
شیر	

نمونه‌ی دیگری که شباهت «ی» و «ن» را با حروف خانواده‌ی «ب» را نشان می‌دهد.

ورودی	خروجی
5 3	1
تا	1
کت	2
کی	
پر	
نر	
یا	
کب	
بر	

## ارسال جواب

برای ثبت نام کاربری و گذرواژه می‌توانید از دستور زیر استفاده کنید که در آن **USERNAME** نام کاربری و **PASSWORD** گذرواژه‌ی پیشنهادی شما هستند.

```
$ echo "register USERNAME PASSWORD" | netcat ct.rudi.ir 40
```

برای ارسال جواب، می‌توانید از دستور زیر استفاده کنید که در آن باید **USERNAME** با نام کاربری، **PASSWORD** با گذرواژه، **LANG** با زبانی که جواب در آن نوشته شده است و **MYPROG** با آدرس فایل جواب جایگزین شوند. به جای عبارت **LANG** مقدار «C»، «C++»، «py»، «java» یا «sh» می‌توان قرار داد که به ترتیب نشان می‌دهند جواب با زبان C، C++، Python، Java یا اسکریپت پوسته نوشته شده است.

```
$ (echo "submit USERNAME PASSWORD ct05 LANG"; \
  cat MYPROG; echo EOF) | netcat ct.rudi.ir 40
```

برای مشاهده‌ی نتایج ارزیابی خودکار جواب‌ها، می‌توانید از دستور زیر استفاده کنید.

```
$ echo "report ct05" | netcat ct.rudi.ir 40
```

ستون‌های خروجی به ترتیب نام کاربری ارسال کننده، زمان ارسال، تعداد نمونه‌ها با خروجی درست و مدت اجرا می‌باشند. برای مشاهده‌ی جواب‌های درست ارسال شده، به ترتیب زمان اجرا، می‌توانید از دستور زیر استفاده کنید.

```
$ echo "report ct05" | netcat ct.rudi.ir 40 | grep '\.$' | sort -nk4
```

در صورت بروز مشکل در اتصال به سرور، در دستورات بالا می‌توانید عبارت «netcat ct.rudi.ir 40» را با عبارت «ssh ct@ct.rudi.ir» جایگزین کنید و وقتی کلمه‌ی عبور از شما درخواست شد، «ct» را وارد نمایید.

## قوانین و نکات

همه‌ی افرادی که در این رقابت شرکت می‌کنند باید قوانین زیر را رعایت نمایند.

- الف** برای حل هر نمونه، به هر جواب دو ثانیه زمان و پانصد مگابایت حافظه اختصاص می‌یابد.
- ب** شرکت‌گندگان آزادند در مورد مسئله‌ها مشورت کنند ولی همه‌ی قسمت‌های جوابی که ارسال می‌شود باید نوشته‌ی شخص ارسال‌کننده باشد.
- ج** ارسال دو جواب مشابه توسط دو کاربر مختلف تخلف محسوب می‌شود.
- د** بین ارسال جواب‌های یک کاربر باید حداقل دو دقیقه فاصله باشد؛ در غیر این صورت، برنامه‌ی داور خطا گزارش می‌دهد.
- ه** استفاده از توازی در جواب‌ها آزاد است؛ می‌توانید از چند پردازنده یا ریسمان استفاده نمایید.
- و** در جلسه‌ای با حضور اعضای محترم تیم فنی، دکتر عمران پور و دکتر غلامی، با توجه به موضوع و سختی مسئله، برای درس‌های مرتبط نمره‌ی اضافی در نظر گرفته شده است.
- ز** در صورت بروز مشکل، آن را در نامه‌ای با عنوانی که با عبارت «CT05» شروع می‌شود به آدرس [gholamirudi@nit.ac.ir](mailto:gholamirudi@nit.ac.ir) گزارش دهید.