



دانشکده مهندسی کامپیوتر

پردازش زبان‌های طبیعی

دکتر بهروز مینایی

زمان تحویل: ۶ آذر

تمرین سری دوم

۱- یک برنامه بنویسید که دو رشته را گرفته و فاصله وزن دار آن‌ها را با سه عمل insert، delete و substitute به دست آورد.

برنامه باید ابتدا وزن هر یک از این سه عمل را گرفته، سپس دو رشته به عنوان ورودی بگیرد و فاصله دو رشته را به صورت وزن دار بدهد. (وزن هر عمل یک عدد طبیعی است).  
همچنین باید برنامه بتواند روش تبدیل رشته اول به رشته دیگر را چاپ کند.

۲- برنامه سؤال یک را طوری اصلاح کنید که از Local alignment نیز پشتیبانی کند.  
برنامه شما باید به طور دقیق شباهت‌های دو رشته و نحوه مطابق شدن آن‌ها را نمایش دهد.

۳- یک برنامه Spelling Correction در زبان انگلیسی پیاده‌سازی کنید.  
برای این کار به چهار جدول نیاز دارید؛ یعنی

- Deletion of Y after X
- Insertion of Y after X
- Substitution of X for Y
- Reversal of XY

تمام این جدول‌ها در پایان مقاله ارسال شده در ضمیمه وجود دارند.

همچنین به یک مجموعه از کلمات زبان انگلیسی به همراه فراوانی آن‌ها در یک پیکره استاندارد نیاز دارید. این مجموعه نیز به نام variances در ضمیمه وجود دارد.

۴- در فصل چهار، یک الگوریتم Dynamic Programming برای پیدا کردن Minimum Edit Distance هنگامی که سه عمل insert، delete و substitute جود دارد، آمده است. حال سعی کنید عمل transposition را نیز به این سه عمل اضافه کنید و یک الگوریتم کارا (چندجمله‌ای) برای پیدا کردن Minimum Edit Distance ارائه کنید. الگوریتم خود را اثبات کرده و پیچیدگی آن را بررسی کنید.

در مورد همه سؤال‌ها تمام فرضیات، نحوه پیاده‌سازی، ایده‌ها و روش‌های شما باید در گزارش وجود داشته باشد.

استفاده از کتابخانه‌های آماده برنامه‌نویسی در این تمرین مجاز نمی‌باشد.

در هر سؤال گزارش خود را در قالب یک فایل pdf آماده کرده و کدها و پیکره‌های تولیدشده خود را همراه با گزارش در یک فایل zip در سایت بارگذاری کنید.

در صورت وجود مشکل، سؤالات خود را به ایمیل‌های [majid.asgari@gmail.com](mailto:majid.asgari@gmail.com) و [y.alizadeh@yahoo.com](mailto:y.alizadeh@yahoo.com) ارسال کنید.