

### تعریف کلی مسئله بدین شرح است:

چند متن فارسی که به زبان عامیانه نوشته شده است و لیستی از کلمات رسمی در اختیار ماست می‌خواهیم متن را به متن فارسی رسمی تبدیل کنیم. اگر کلمه ای در متن است که در لیست کلمات ذکر شده است آن کلمه رسمی است در غیر این صورت کلمه ای عامیانه است بنابراین تصمیم می‌گیریم آن را با یکی از کلمات رسمی که در اختیار داریم جایگزین کنیم.

### این کلمه جایگزین کدام کلمه لیست است؟

✓ کلمه ای که به کلمه عامیانه شباهت بیشتری دارد یا به عبارتی دیگر فاصله کمتری با آن دارد.

### فاصله کلمات لیست را از کلمات عامیانه چگونه محاسبه نماییم؟

- ✓ با استفاده از معیاری به نام Edit Distance
- ✓ یا با استفاده از معیاری به نام Longest Common Subsequence

▪ نکته: یکی از این روش ها باید انتخاب شود.

### تعریف مسئله محاسبه Edit-distance:

برای این کار، فرض می‌کنیم، می‌خواهیم کلمه غلط (a) را به هر کلمه لیست (b) تبدیل نماییم. برای این تبدیل، مجاز به انجام 3 حرکت زیر هستیم:

• حذف یک حرف

• اضافه کردن یک حرف

• تغییر یک حرف

کمترین تعداد ممکن تغییرات برای تبدیل a به b، فاصله کلمه a از b است. این فاصله را می‌توانید با استفاده از تکنیک dynamic programming در زمان قابل قبولی محاسبه کنید.

فاصله a از تمام کلمات لیست را می‌یابیم و چند کلمه با کمترین فاصله را به عنوان پیشنهاد اعلام می‌نماییم.

در این پروژه، می‌خواهیم ابتدا مسئله محاسبه Edit-distance را با استفاده از تکنیک dynamic programming حل کرده و سپس راه حل پیشنهادی را در کاربرد spell-checking پیاده‌سازی نماییم.

## مرحله بندی پروژه:

در مرحله اول پروژه یا همان حل مسئله، باید یک گزارش که حاوی مطالب زیر است تحویل داده شود.

- تعریف مسئله با ذکر مثال
  - تعریف فرم بهینه بازگشتی
  - رسم درخت بازگشت و نمایش تکرار مسائله
  - ارائه راه حل پویا (پایین به بالا) و محاسبه پیچیدگی آن
- مرحله دوم پروژه یا همان مرحله پیاده سازی شامل مراحل زیر است. برای پیاده سازی کد لازم است از یک دیکشنری استفاده نمایید. قصد ما اجرای این برنامه زبان فارسی است. برای این منظور یک دیکشنری زبان فارسی در اختیار شما است که ضمیمه همین تمرین است.
- تعیین کلمات عامیانه در متن
  - محاسبه فاصله کلمه عامیانه از کلمات لیست با استفاده از راه حل پویای فوق
  - پیشنهاد نزدیکترین کلمات
  - محاسبه مرتبه زمانی الگوریتمتان
- دو مرحله اول کافیهست به درستی انجام شوند و شما نمره آن را دریافت خواهید کرد. اما هرچه سریعتر باشند (به صورت رقابتی) نمره اضافه بیشتری دریافت خواهید نمود. سریع بودن الگوریتمتان را باید توجیه کنید.

## ویژگی های امتیازی پروژه:

- ❖ رابط کاربری گرافیکی (GUI)
- ❖ اضافه کردن الگوریتمی جدید برای محاسبه نزدیک ترین کلمه به الگوریتم انتخابی قبلی
- ❖ نمایش گرافیکی پیچیدگی زمانی الگوریتم محاسبه نزدیک ترین کلمه
- ❖ استفاده از تمام لغت های فایل txt

## نکات:

- پروژه در گروه های حداکثر سه نفره انجام می شود.
- ارائه پروژه به صورت حضوری انجام می شود و تمامی اعضا باید تسلط کامل داشته باشند.
- در صورت وجود تشابه نمره پروژه هر دو گروه صفر می شود.