

Full Text Search

問題：野田

解答作成：北川・平澤

スライド作成：平澤

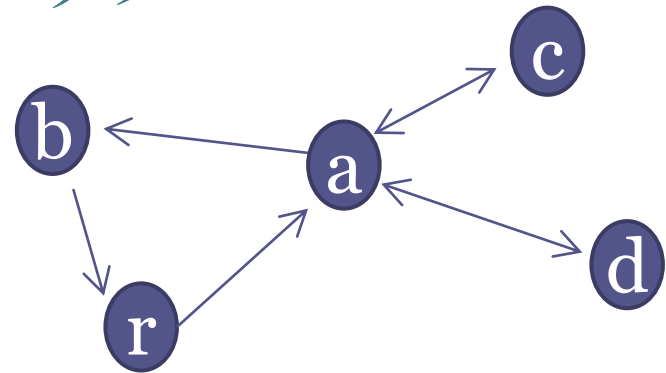
問題概要

- 与えられた検索クエリ Q に対し、その検索クエリに含まれる **1-gram** と **2-gram** を全て含んだような文字列の最小の長さを返せ
- ただし、 Q 自体を含まない文字列であること
 - アルゴリズムが直ぐに思いつかないときは、いろいろ例を考えて、試してみましょう

初めにやること：入力を変換

- 2-gramまでが考慮される
 - それ以外の情報はもはや不要
 - 各文字をノードとするグラフ

- 例: abracadabra -->



- グラフ上の全ての枝を通り、かつ元の文字列を含まないもので、最短の順路
 - グラフ上にない枝を追加してもOK

“元の文字列を含まない”？

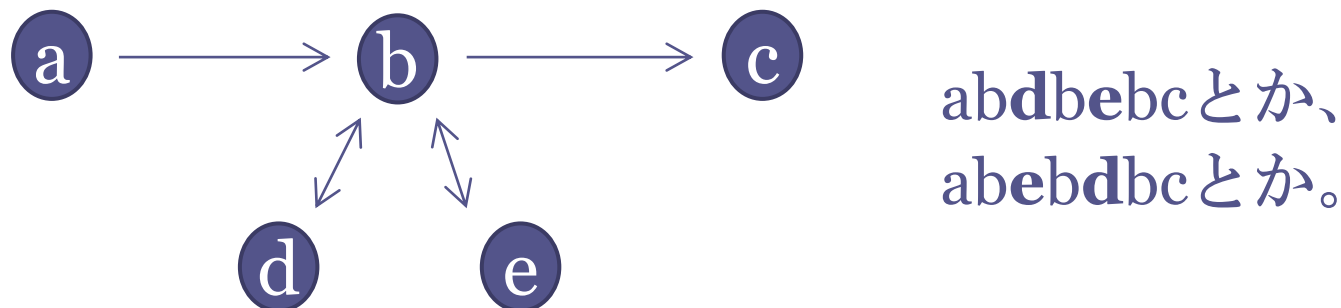
- クエリ中に2回以上出てくる2-gramがあれば一部を消すことで元の文字列から変えられる
 - 例： abracadabra -> abracada
- 元の文字列と同じ長さになるなら2通り以上の作り方があれば良い
 - 例： abracada -> acadabra
- それも無理なら1文字増やしてやる
 - 例： abr -> abbr / brab 等

“2通り以上の作り方”？

- 例えば以下のようなとき
 - 全体がオイラー路になっている



- オイラーパス中にいろいろな順路がある



まとめ

1. Queryが2文字以下なら No Results を出す
2. 入力をグラフ上に落としてやる
3. グラフの特徴に応じて場合分け
 1. オイラー路になっていれば？
 2. オイラーパスになっていれば
 1. クエリ中に同じ2-gramが複数回入っている？
 2. クエリ中に同じ2-gramが現れなければ
 1. パスの作り方が1通りなら？
 2. パスの作り方が複数あれば？
 3. それ以外なら？

考えてみてください