

# Full Text Search

問題：野田

解答作成：北川・平澤

英文作成：平澤

英文校正：泉・稲葉・吉田

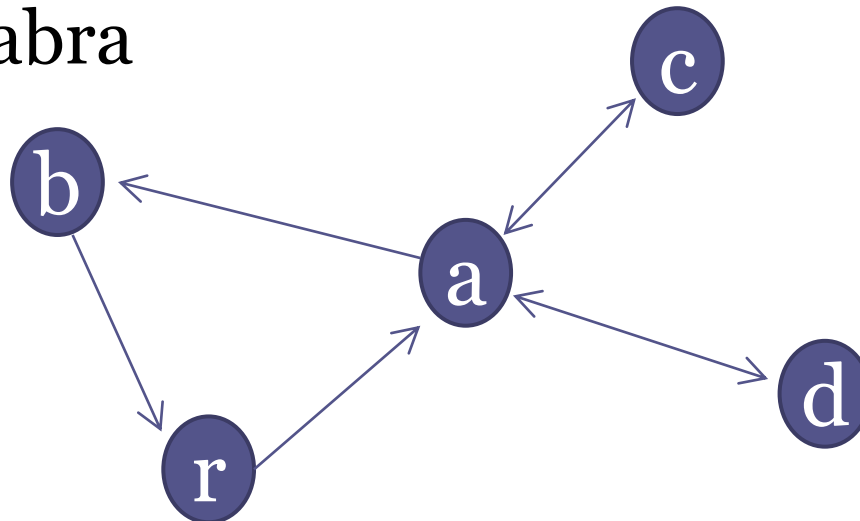
スライド作成：平澤

# 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

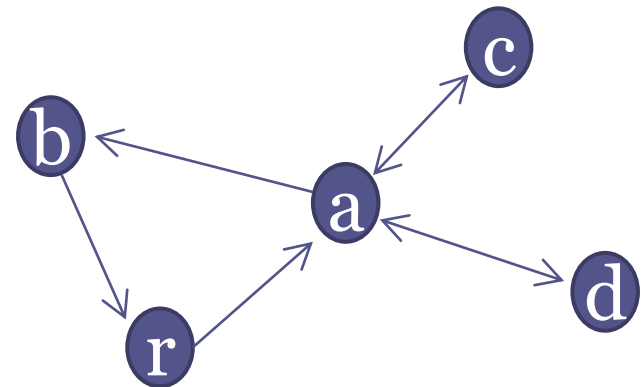
## 初めにやること：入力を変換

- 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が良さそう
- 例: abracadabra



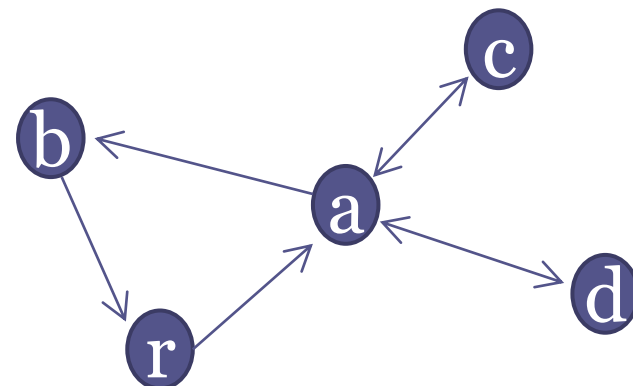
# 入力を変換した結果

- 以下  $3 \leq \text{len}(\text{str})$  を仮定
  - 1-gramは2-gramに含まれる
- グラフの枝: 2-gramに対応
- 以下の問題を解けばいい
  - 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、最短の経路長
    - グラフ上にない枝を追加してもOK



# 入力を変換した結果

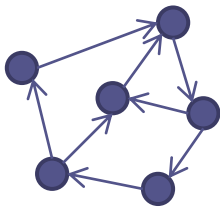
- 以下  $3 \leq \text{len}(\text{str})$  を仮定
  - 1-gramは2-gramに含まれる
- グラフの枝: 2-gramに対応
- 以下の問題を解けばいい
  - グラフ上の全ての枝を通り、かつ元の文字列を含まないもので、最短の経路長
    - グラフ上にない枝を追加してもOK



→問題が複雑なときはもっと簡単にしてから考えてみよう

# 全ての枝を通る、最短の経路長は？

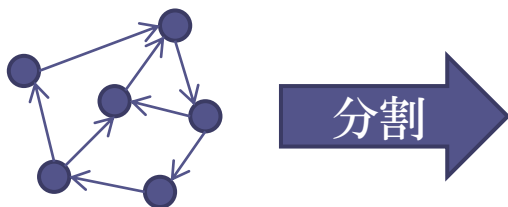
- 元の文字列を含んでも良いことにする。
- 有向オイラー(閉)路になっていれば
  - その有向オイラー(閉)路の長さ
- otherwise:
  1. グラフ上の枝を分割して、複数の路の集合にする
  2. グラフ上にその路をつなぐ枝を加えた、その長さ



- 分割数が最小なら最終経路長も最小

## 全ての枝を通る、最短の経路長は？

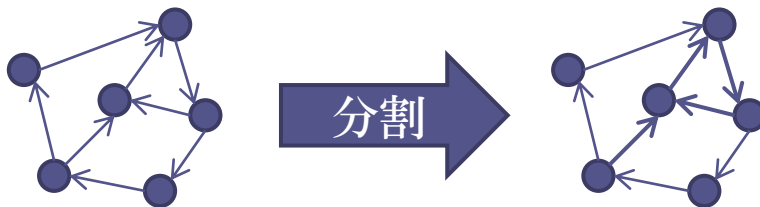
- 元の文字列を含んでも良いことにする。
- 有向オイラー(閉)路になっていれば
  - その有向オイラー(閉)路の長さ
- otherwise:
  1. グラフ上の枝を分割して、複数の路の集合にする
  2. グラフ上にその路をつなぐ枝を加えた、その長さ



- 分割数が最小なら最終経路長も最小

# 全ての枝を通る、最短の経路長は？

- 元の文字列を含んでも良いことにする。
- 有向オイラー(閉)路になっていれば
  - その有向オイラー(閉)路の長さ
- otherwise:
  1. グラフ上の枝を分割して、複数の路の集合にする
  2. グラフ上にその路をつなぐ枝を加えた、その長さ

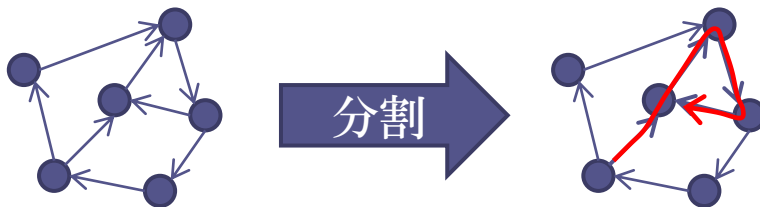


- 分割数が最小なら最終経路長も最小



# 全ての枝を通る、最短の経路長は？

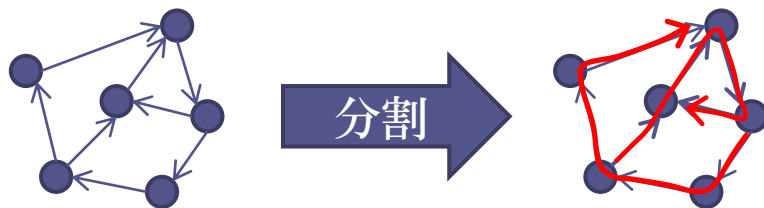
- 元の文字列を含んでも良いことにする。
- 有向オイラー(閉)路になっていれば
  - その有向オイラー(閉)路の長さ
- otherwise:
  1. グラフ上の枝を分割して、複数の路の集合にする
  2. グラフ上にその路をつなぐ枝を加えた、その長さ



- 分割数が最小なら最終経路長も最小

## 全ての枝を通る、最短の経路長は？

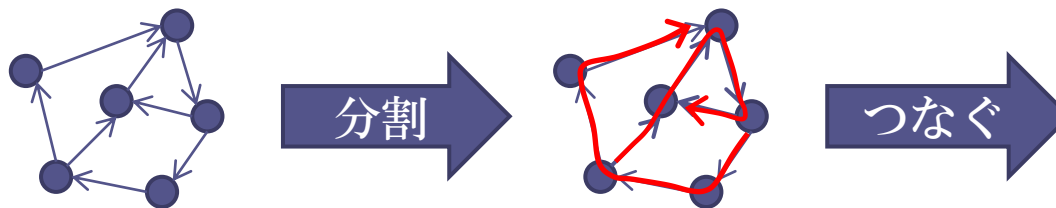
- 元の文字列を含んでも良いことにする。
- 有向オイラー(閉)路になっていれば
  - その有向オイラー(閉)路の長さ
- otherwise:
  1. グラフ上の枝を分割して、複数の路の集合にする
  2. グラフ上にその路をつなぐ枝を加えた、その長さ



- 分割数が最小なら最終経路長も最小

# 全ての枝を通る、最短の経路長は？

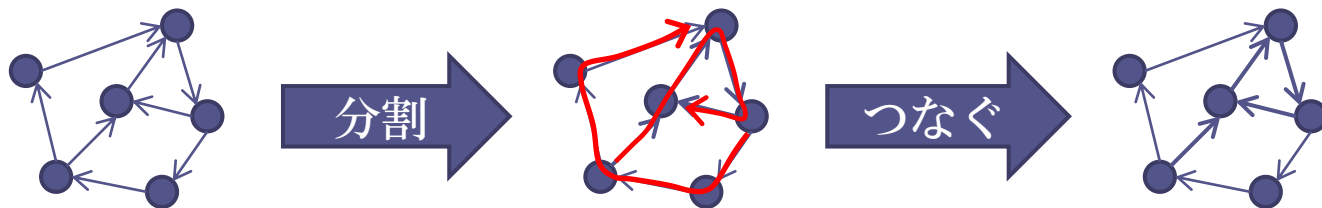
- 元の文字列を含んでも良いことにする。
- 有向オイラー(閉)路になっていれば
  - その有向オイラー(閉)路の長さ
- otherwise:
  1. グラフ上の枝を分割して、複数の路の集合にする
  2. グラフ上にその路をつなぐ枝を加えた、その長さ



- 分割数が最小なら最終経路長も最小

# 全ての枝を通る、最短の経路長は？

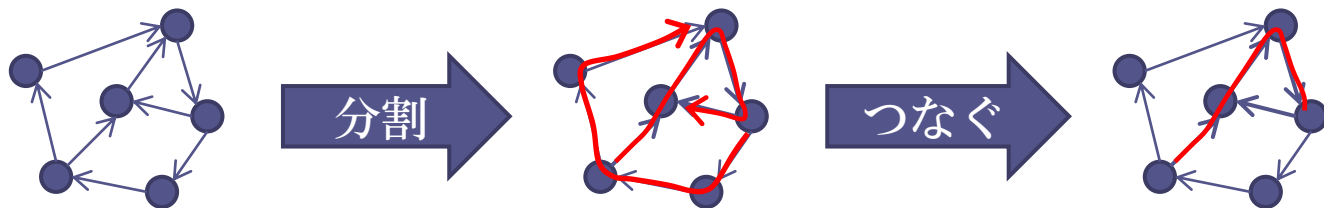
- 元の文字列を含んでも良いことにする。
- 有向オイラー(閉)路になっていれば
  - その有向オイラー(閉)路の長さ
- otherwise:
  1. グラフ上の枝を分割して、複数の路の集合にする
  2. グラフ上にその路をつなぐ枝を加えた、その長さ



- 分割数が最小なら最終経路長も最小

# 全ての枝を通る、最短の経路長は？

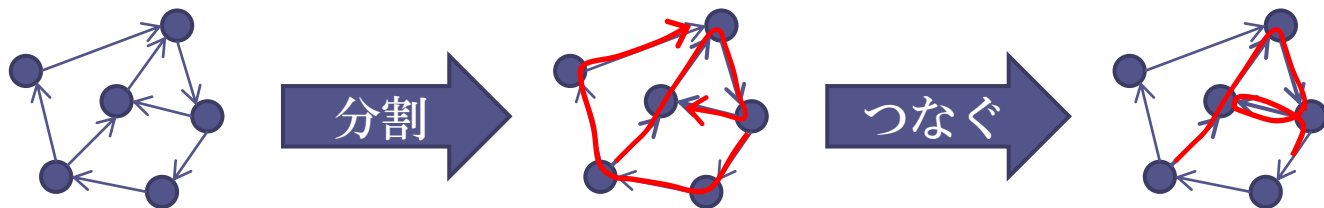
- 元の文字列を含んでも良いことにする。
- 有向オイラー(閉)路になっていれば
  - その有向オイラー(閉)路の長さ
- otherwise:
  1. グラフ上の枝を分割して、複数の路の集合にする
  2. グラフ上にその路をつなぐ枝を加えた、その長さ



- 分割数が最小なら最終経路長も最小

# 全ての枝を通る、最短の経路長は？

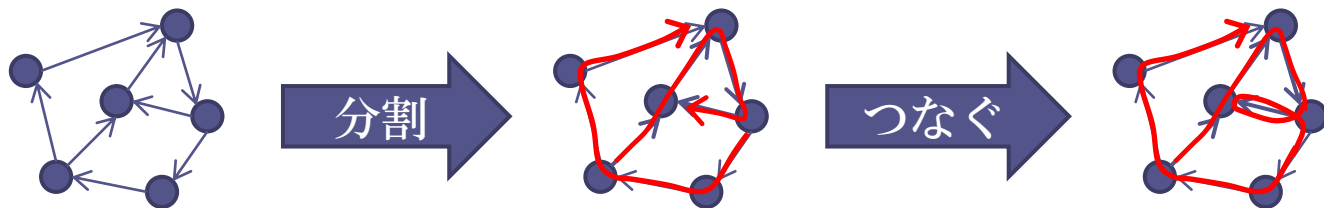
- 元の文字列を含んでも良いことにする。
- 有向オイラー(閉)路になっていれば
  - その有向オイラー(閉)路の長さ
- otherwise:
  1. グラフ上の枝を分割して、複数の路の集合にする
  2. グラフ上にその路をつなぐ枝を加えた、その長さ



- 分割数が最小なら最終経路長も最小

# 全ての枝を通る、最短の経路長は？

- 元の文字列を含んでも良いことにする。
- 有向オイラー(閉)路になっていれば
  - その有向オイラー(閉)路の長さ
- otherwise:
  1. グラフ上の枝を分割して、複数の路の集合にする
  2. グラフ上にその路をつなぐ枝を加えた、その長さ



- 分割数が最小なら最終経路長も最小

# 枝の最小分割数を求める

- 実は簡単
- 最小分割数 =  
各ノードの $\max(\text{出次数} - \text{入次数}, 0)$ の和
  - c.f.) 有向オイラー路：
    - 連結かつ各ノードの $\max(\text{出次数} - \text{入次数}, 0)$ の和  $\leq 1$
  - c.f.) 有向オイラー閉路 (上の式の唯一の例外)：
    - 連結かつ各ノードの $\max(\text{出次数} - \text{入次数}, 0)$ の和  $= 0$



# 枝の最小分割数を求める

- 実は簡単
- 最小分割数 =  
各ノードの $\max(\text{出次数} - \text{入次数}, 0)$ の和
  - c.f.) 有向オイラー路：
    - 連結かつ各ノードの $\max(\text{出次数} - \text{入次数}, 0)$ の和  $\leq 1$
  - c.f.) 有向オイラー閉路 (上の式の唯一の例外)：
    - 連結かつ各ノードの $\max(\text{出次数} - \text{入次数}, 0)$ の和  $= 0$

証明略；そんなに大変ではないです。

## ここまでのまとめ

- 元の文字列を含んでもよい場合
  - 長さが3以上の場合を考えています
- 全体がオイラー(閉)路のとき
  - ⇔各ノードの $\max(\text{出次数}-\text{入次数}, 0)$ の和 $\leq 1$
  - オイラー(閉)路の長さ(グラフ上の枝の数+1)が解
- Otherwise
  - グラフ上の枝の数+  
各ノードの $\max(\text{出次数}-\text{入次数}, 0)$ の和

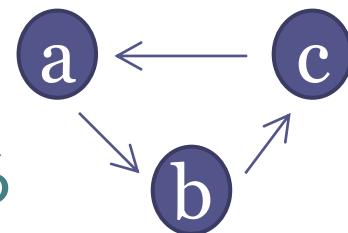
## ここまでのまとめ

- 元の文字列を含んでもよい場合
  - 長さが3以上の場合を考えています
- 全体がオイラー(閉)路のとき
  - ⇔各ノードの $\max(\text{出次数}-\text{入次数}, 0)$ の和 $\leq 1$
  - オイラー(閉)路の長さ(グラフ上の枝の数+1)が解
- Otherwise
  - グラフ上の枝の数+  
各ノードの $\max(\text{出次数}-\text{入次数}, 0)$ の和

→簡単な問題が解けたので、元々の問題を考えてみよう

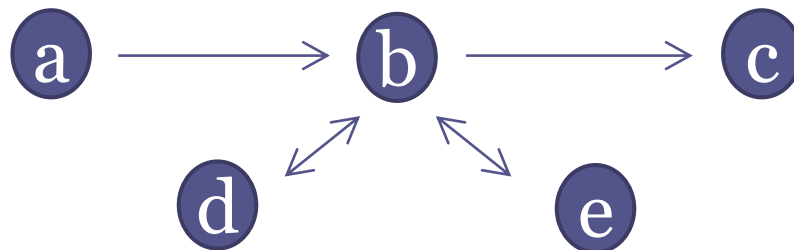
# 元の文字列を含んではいけない時

- まず、長さが2以下なら No Results
  - 以下、長さが3以上の場合を考える
- 同じ長さで複数の文字列が作れるなら問題ない。
- 全体がオイラー閉路のとき
  - 開始点を変えることで複数の文字列が作れる
- 全体がオイラー路のとき(閉路を除く)
  - 状況により異なるので次スライドで。
- Otherwise
  - 路をつなぐ順序を変えることで複数の文字列が作れる



## 全体がオイラー路のとき (閉路を除く)

- オイラー路の長さ < 元の文字列の長さ なら
  - これをそのまま出力すればok
- Otherwise (上記不等号が=のとき)
  - 同じ長さで複数の文字列が作れるか判定する
    - ・ 少し難しいですが挑戦してみましょう
- 例) 同じ長さで複数の文字列が作れる場合



abdbecとか、  
abebdbcとか。

# 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. オイラー閉路になっていれば？
  2. オイラー路になっていれば (閉路を除く)
    1. オイラー路の長さ<クエリの長さなら？
    2. オイラー路の長さ=クエリの長さなら？
      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？

## 出典 / ジャツジ解

- 実際にあった話から作題されたようです
  - QMAClone BugTrack
    - <http://kishibe.dyndns.tv/qmaclone/wiki/wiki.cgi?page=BugTrack-QMAClone%2F359>
- ジャツジ解
  - 北川      93行    1719B
  - 平澤    113行   2189B

# 結果

- Submitチーム数： 5
- Acceptチーム数： 4
- 総Submit数： 11
  
- First Submit & FirstAccept:
  - 2h 57m
  - m(\_\_\_)m (National Taiwan University)