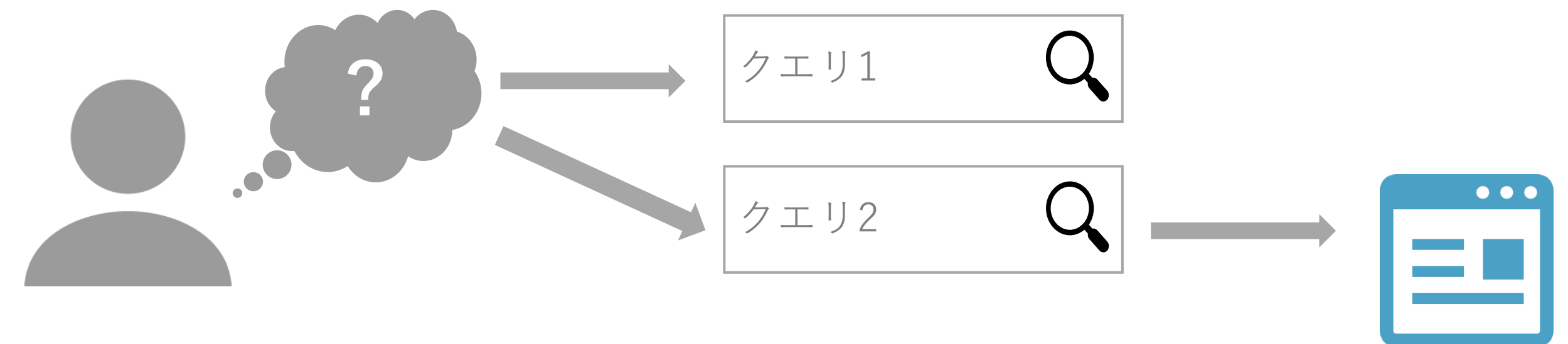


背景・目的：検索クエリ作成時のユーザの工夫の解明

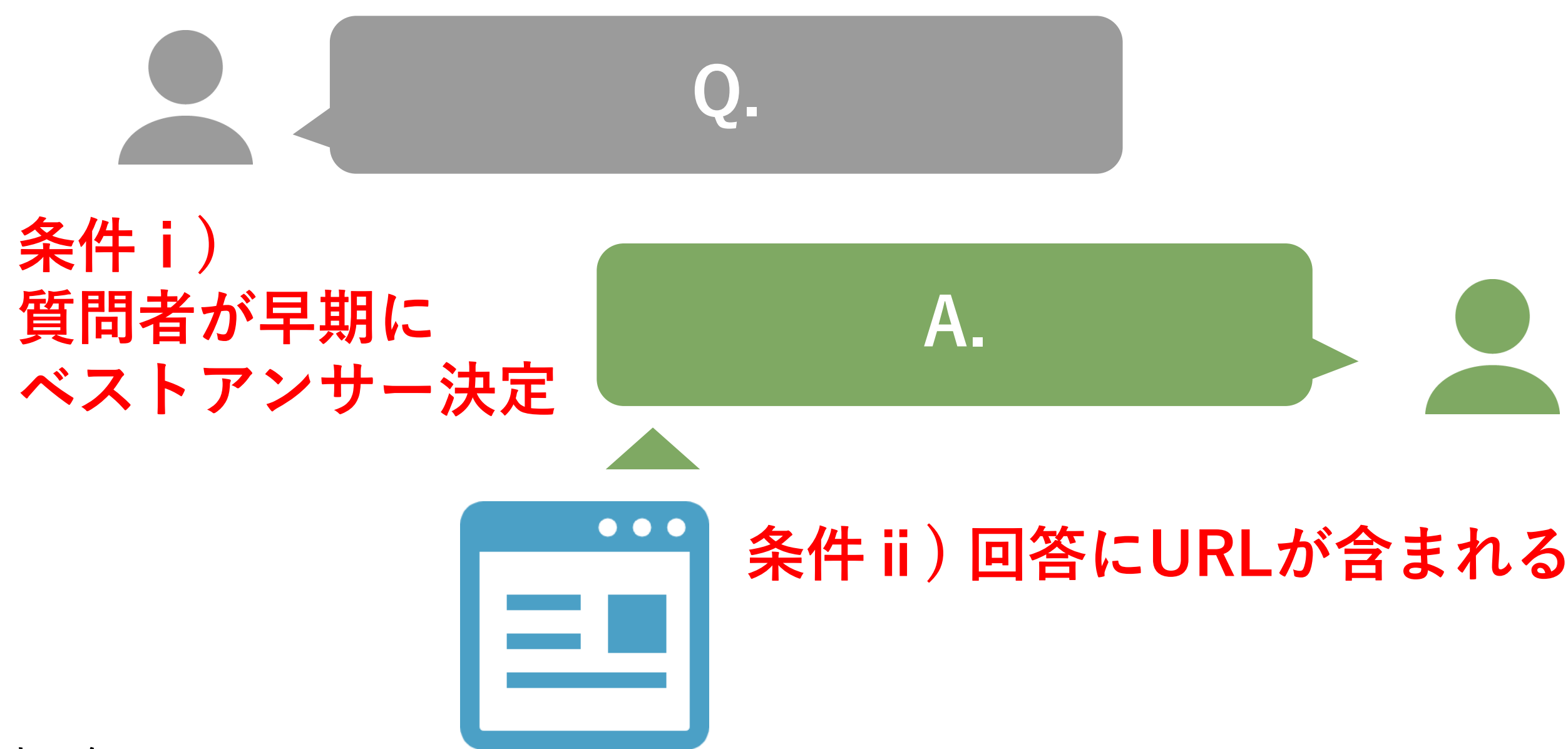
- ユーザは、情報ニーズを、**望む情報がより得られそうな検索クエリに変換している**
- その変換の工夫を解明することで、**情報検索の支援につなげられないか？**



提案手法：Yahoo!知恵袋から、質問・その回答となるwebページを取り出す

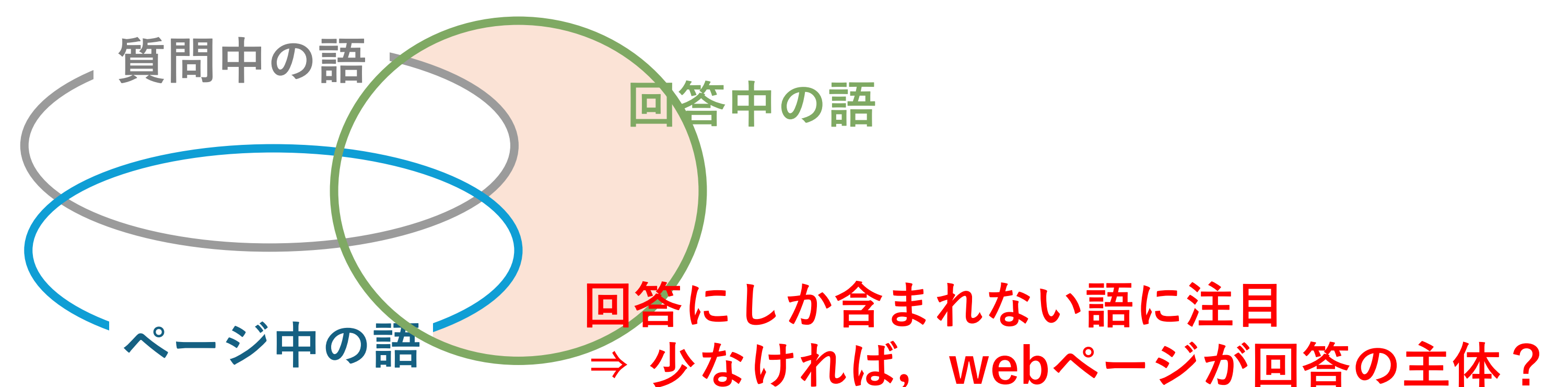
- ヒューリスティックに設定した条件に基づく絞り込みのうえで、回答に含まれるwebページだけでも回答として成立しうるかを調べた

① リンク付き・高満足回答の絞り込み



② 回答におけるリンク先ページの重み調査

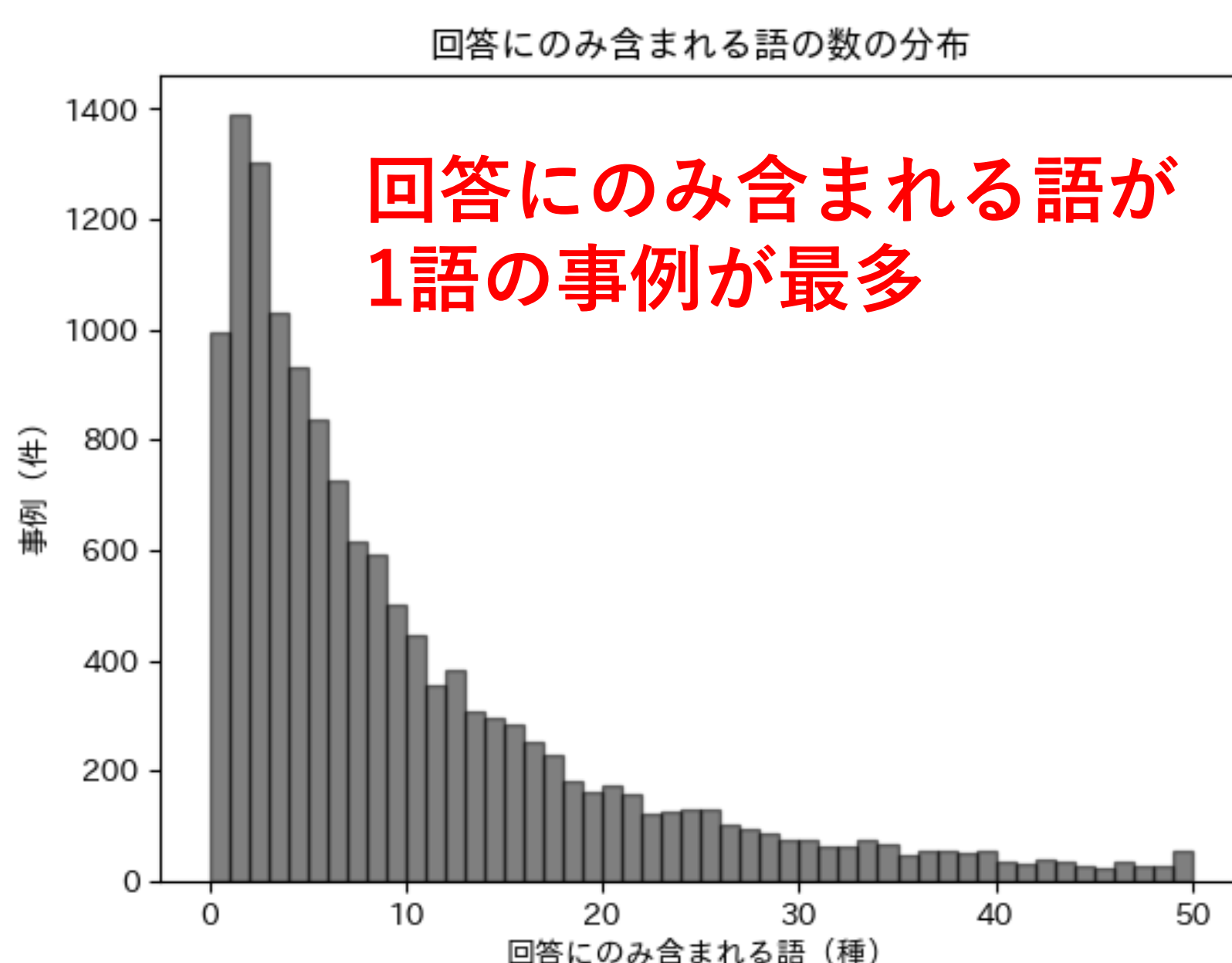
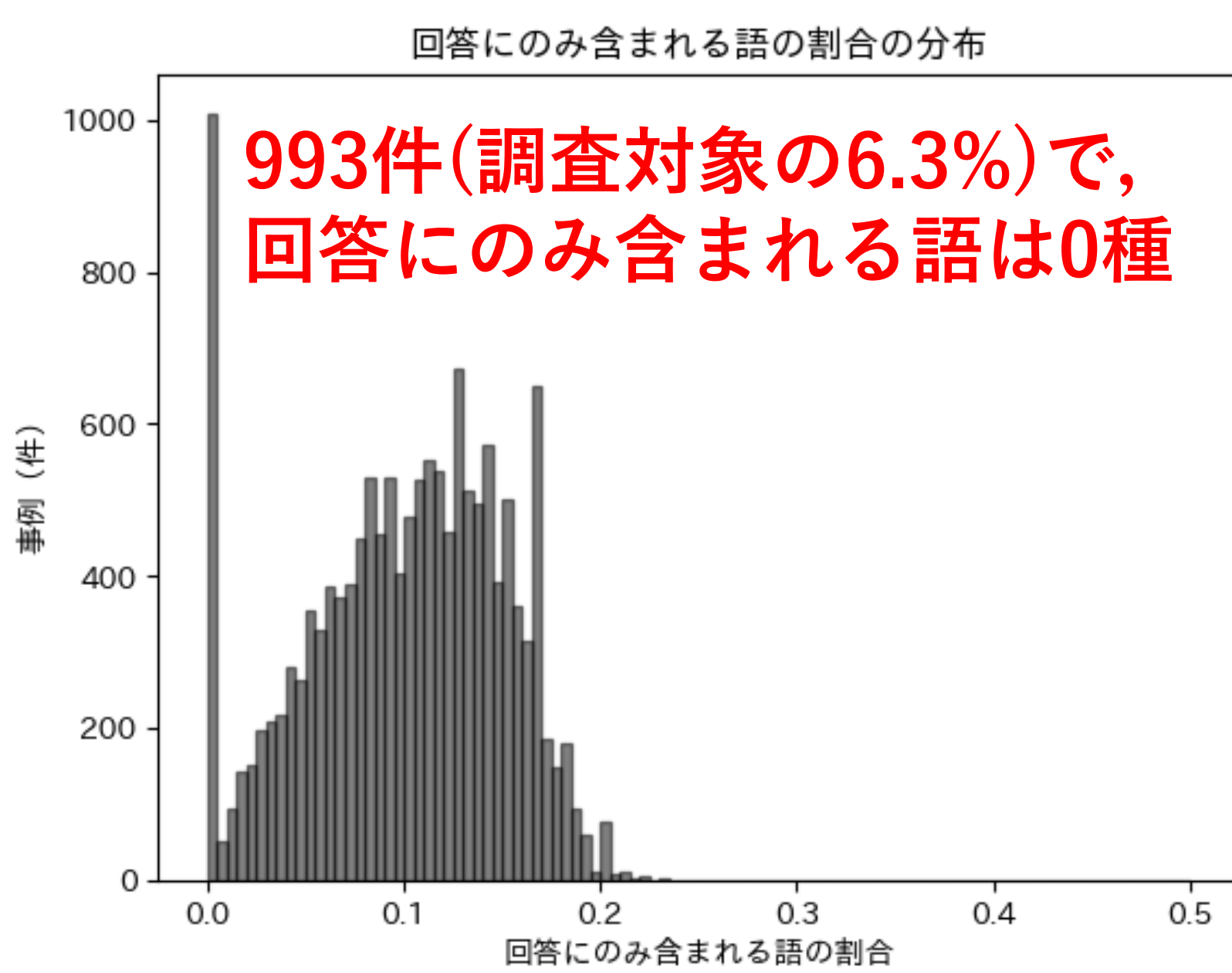
- 質問・回答・リンク先ページ(本文のみ)を形態素解析して、それぞれに含まれる語の集合を作成



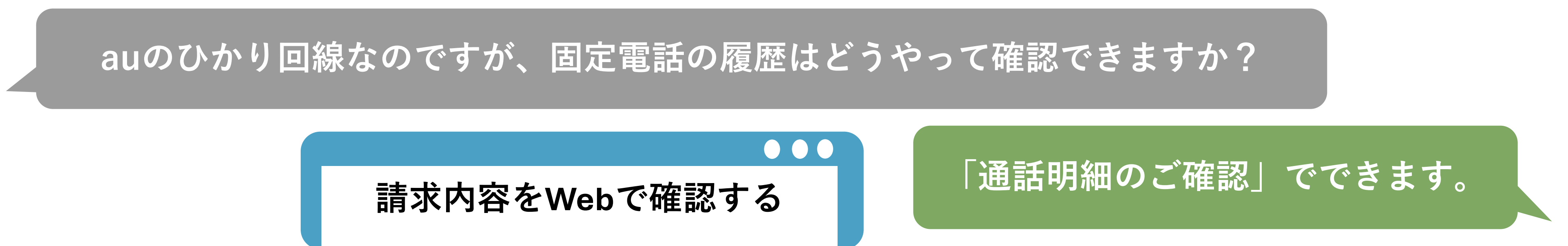
- 実験
 - 簡単のため、(iii)回答・回答中のURLが1つ (iv)動画画像を含まない (v)ページの入手/解析が可能の条件も満たす**1.5万件**を対象として、
 - 手順②を実施した

結果・考察：回答者の検索の工夫が見える事例を発見可能

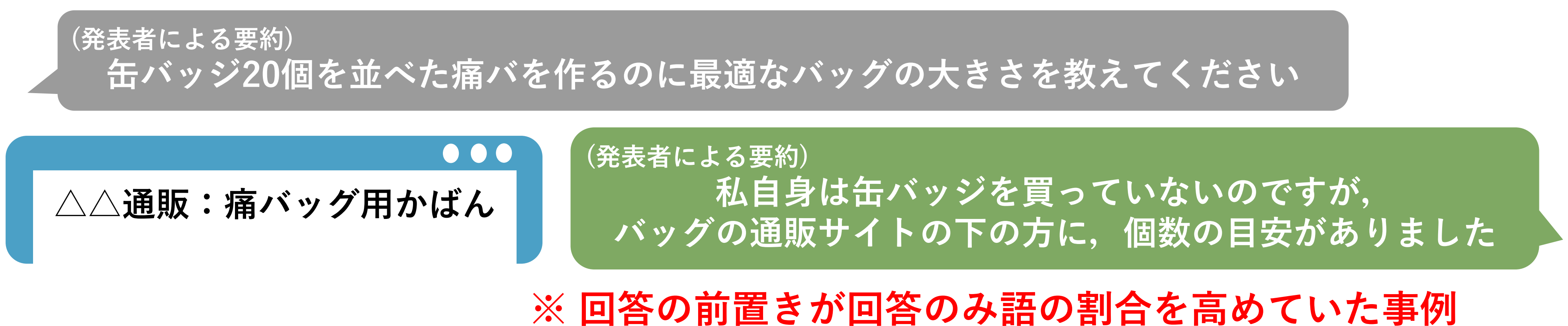
- 回答がほぼwebページの情報である事例の存在が確認できた



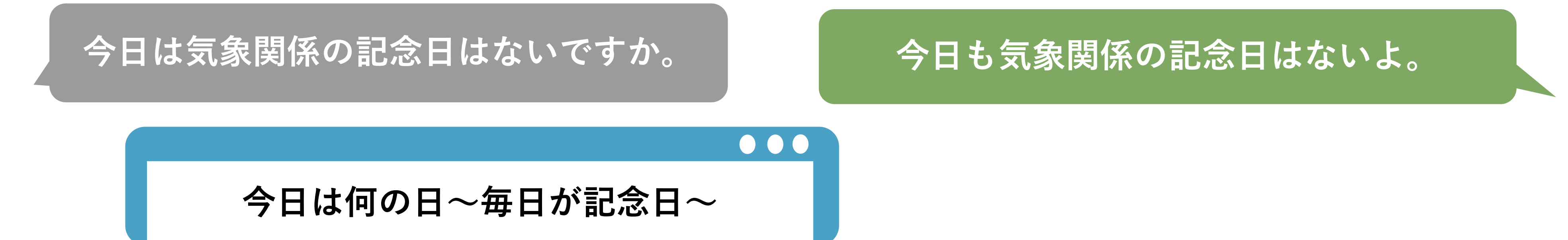
- 事例1：適切な用語への言い換え (回答のみ語 0)



- 事例2：やや主題のずれたページの活用 (回答のみ語 45, 回答のみ語割合11%)



- 事例3：特定領域の専門ページの利用 (回答のみ語 0)



展望

(1) 手法の改善

- 回答にのみ含まれるが、webページ以上に質問に答えてはいない語の判別・除外
- 再現率・精度の最適点の模索

(2) 取り出された質問・webページを用いた情報検索行動研究

- 回答者の提示したページは、質問者の用いたキーワードではヒットしないか？
- 質問者の用いたキーワードと、回答者の提示したページのキーワードの意味的な関係は？