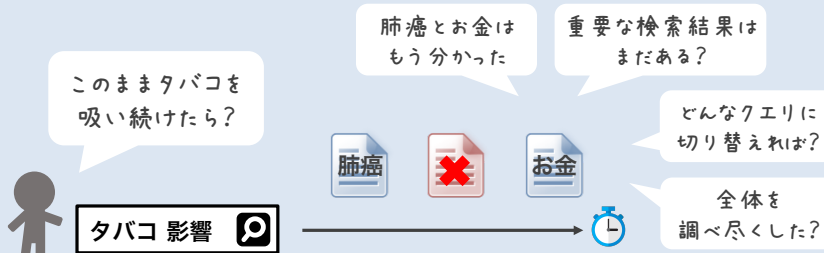


網羅性指向タスクにおける未閲覧情報量の提示インターフェース

網羅的信息収集タスク

トピックの多様な観点に関する情報の網羅的な収集が必要



問題① 適切なクエリの作成が困難



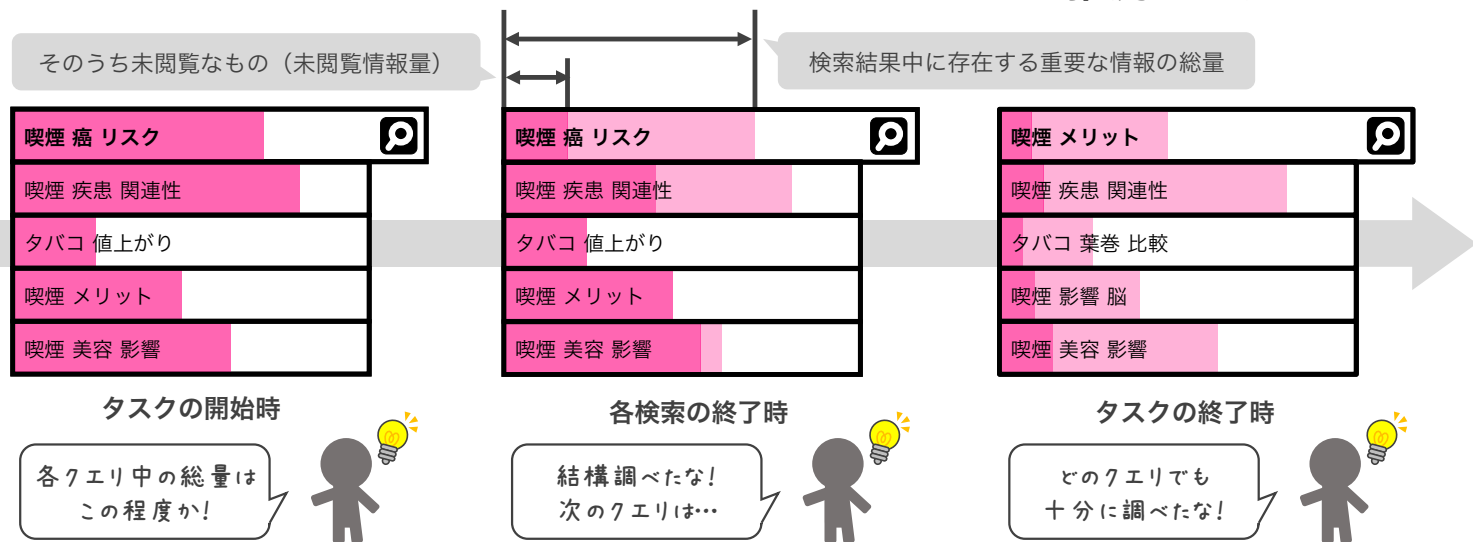
網羅性の低い検索結果 → 検索回数の増加

問題② 検索の終了判断が困難



長期継続 → 時間の無駄 早期終了 → 重要情報の見逃し

提案 これまでの検索で調べ切れていない重要な情報の量を可視化するクエリ推薦インターフェース



定式化

t : 検索トピック
 D_u : ユーザ u の既閲覧文書集合,
 D_q^K : クエリ q の上位 K 件の検索結果文書集合

未閲覧情報量 = 検索結果中の未閲覧文書集合の閲覧により獲得可能な追加の利得

$$MI_{u,t}(q) = Gain_t(D_u \cup D_q^K) - Gain_t(D_u)$$

既閲覧文書と検索結果文書の両方から得られる利得

既閲覧文書から得られた利得

(サブトピックマイニング手法を利用して値を推定)

利得 3種類の性質を考慮

重要性: 重要な観点に関する文書 > 瓊末な観点に関する文書

$$Gain_t(D) = \sum_{a \in A_t} Pr(a|t) \cdot Gain_a(D)$$

適合性: 観点適合度の高い文書 > 観点適合度の低い文書

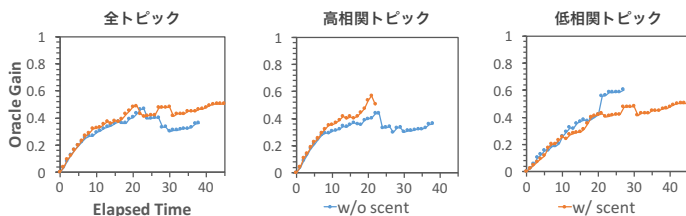
$$Gain_a(D) = \sum_{i=1}^{|D|} Rel_a(d_i) \cdot Disc_a(\{d_1, \dots, d_{i-1}\})$$

新規性: 未探索の観点に関する文書 > 探索済の観点に関する文書

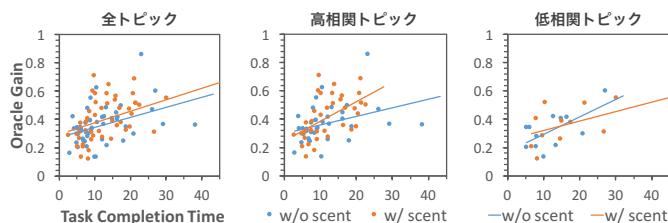
$$Disc_a(D') = \prod_{a'_j \in D'} (1 - Rel_a(d'_j))$$

- 評価**
- 24名の被験者・4種類のトピックでユーザ実験: 提案UI (w/ scent) vs. 未閲覧情報量なしUI (w/o scent)
 - NTCIR INTENT, INTENT-2, Imine タスクのトピックと正解データを利用

結果#1 提案UI利用者は一定時間経過後も高い利得を獲得



結果#2 提案UI利用者は単位時間当たりに高い利得を獲得



推定精度の高い未閲覧情報量の提示 → 網羅的信息収集タスクにおけるユーザの検索行動を (非) 効率化