

NAL研紹介資料その2

目次

- ・ 動機に繋がってる事例集

- 人工頭脳プロジェクト(ロボットは東大に入れるか)
- ANPI_NLP(東日本大震災のためのデータマイニング・自然言語処理)

- ・ 具体的にやってる/やりたいテーマ

- コンテキスト推定
- 議論構造獲得
- 発言間関係の例
- NAL研テーマ
 - ・ 仮想秘書
 - ・ 子どものように育てられるシステム

- ・ (NAL研の歩み)

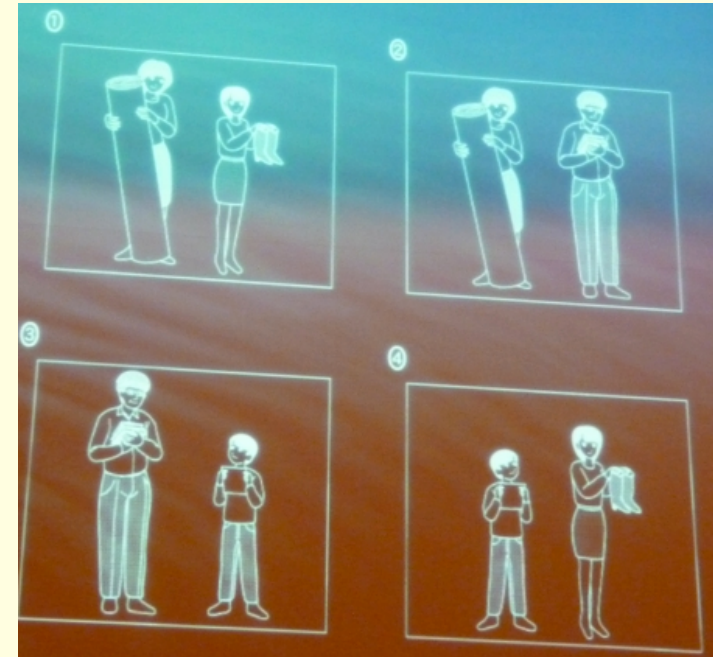
〈動機に繋がっている〉 事例集

～NAL研～

「イラストによって表象され、それを理解できる」には何が必要か？

- 問題の例(英語)
 - ある一家の一日。父と娘に関する語りを説明。
 - それに相応しいイラストを選択する問題。
- 問題の特性/特質
 - 膨大なデータがあれば解決できる？
 - 知識／常識／演繹の中でも「常識」が厄介。
 - 【身体性が持つある種の感性や感覚、それが醸し出す常識が渾然一体となっている】

- 科学に提供されている言語：数学の言語（解析、代数、確率統計、論理、確率統計、、）
- この言語で書けるならコンピュータに搭載できる。
- 逆に書けないなら、搭載できない。
- 常識や身体性から来る感覚をうまく記述する方法を見つける必要がある。



<http://bit.ly/AAlWuY> より転載

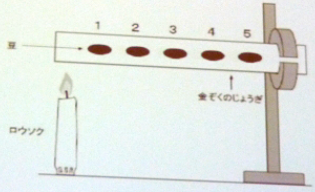
出典：「人工頭脳プロジェクト(ロボットは東大に入れるか)」キックオフシンポジウム (2011/12/14)、プロジェクト概要説明 (新井紀子・NII)

「実世界を理解する」には何が必要か？

- 国際数学・理科教育動向調査 (TIMSS)
 - 定規にバターで固定した豆を用意。
 - 片方から温めるとどの順番で豆が落ちるか。
 - 小学4年理科：正解率日本92%、世界57%
 - 温めるとバターが溶ける／温められる場所から／etc.
 - バターの融点や重力の方程式を計算していても解けないが、常識では簡単に解ける。
- 実体験に基づく論理的思考が必要な問題の例
 - 自転車で転び、袋に入れた食塩をこぼしてしまった。
 - 集めてビニール袋に戻したが砂や落ち葉も混ざった。
 - 食塩だけを取り出す方法を順番に書き、その理由も書きなさい。
 - 教科書に書かれていることの理解には「実体験との接続」が重要。

二つの流れが合流して初めて解ける問題の例

- 金属のじょうぎにバターで豆をくっつけた。じょうぎの一方の端を暖めた場合、豆はどの順番で落ちるか？
- 国際数学・理科教育動向調査 (TIMSS) 2007
- 小学4年理科：正解率日本92%、世界57%




<http://www.nier.go.jp/timss/2007/index.html>

<http://bit.ly/AAlWuY> より転載

実体験に基づく論理的思考が必要な問題の例

- 太郎さんは自転車で転び、持っていた袋の中に入っていた食塩をこぼしてしまいました。彼はこぼれた食塩を集めてビニール袋に戻しましたが、一緒に砂や落ち葉も混ざってしまいました。袋の中から食塩だけを取り出す方法を順番に下の表に書き込みなさい。またその理由も書きなさい (一部略)



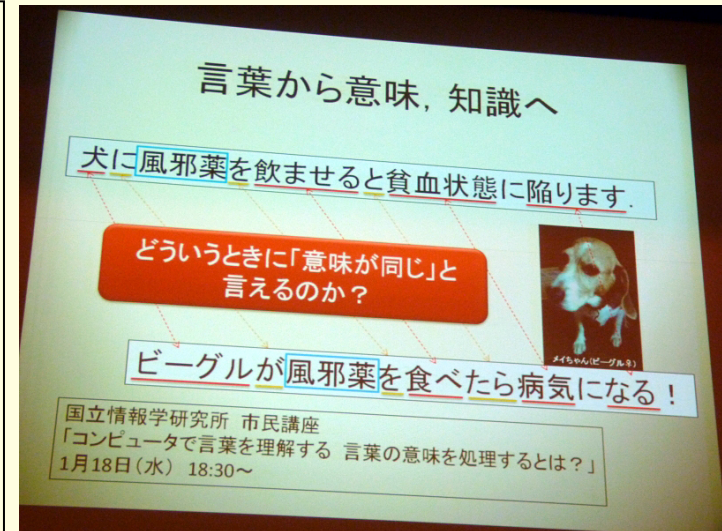
順番	方法の説明	その理由
1	中身をふるいにかける	落ち葉を除去できる
2		
3		
4		

国際数学・理科教育動向調査 (TIMSS) 2007 中学2年
<http://www.nier.go.jp/timss/2007/index.html>

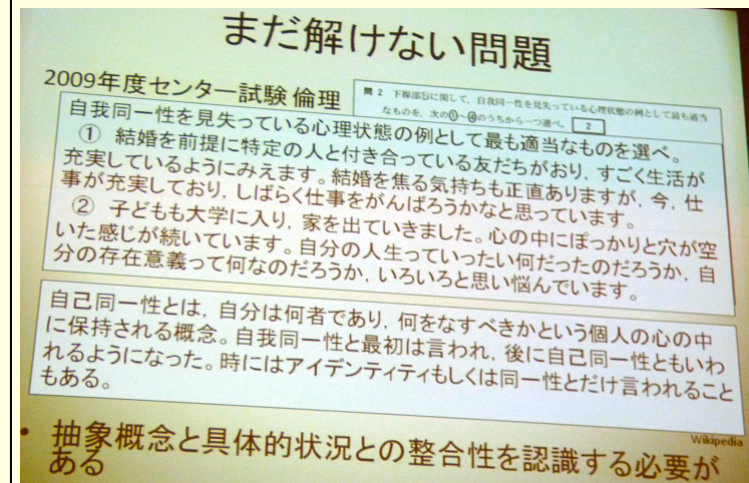
出典：「人工頭脳プロジェクト(ロボットは東大に入れるか)」キックオフシンポジウム (2011/12/14)、人工知能にとってなぜ大学入試はチャレンジングか (稲邑哲也・NII)

「意味的に一致しているかを認識する」 には何が必要か？

- 知識を問う問題とは？
 - 記憶していることと問われていることが「意味的に一致しているかどうか」を認識する。
- 含意関係認識
 - 2つの文の間に含意関係が成り立つかどうかを判定。
 - 含意関係：ある文が正しいとした時、もう一方の文が正しいと言える。
- 含意関係でできることはごく一部
- 「知識を問う問題」の先へ
 - 自然言語→意味・知識
 - 出題意図／含意関係／知識検索
 - シミュレーション／数式処理・理解／知識統合／時間・空間の推論／常識／因果関係の認識・推論／etc.
- その先には何があるのか。



<http://bit.ly/AAlWuY> より転載



出典: 「人工頭脳プロジェクト(ロボットは東大に入れるか)」キックオフシンポジウム
(2011/12/14)、知識を問う問題にコンピュータはどれだけ答えられるか (宮尾祐介・NII)
研究室紹介(NAL研), 2013.10.30 <http://www.eva.ie.u-ryukyu.ac.jp/~tnal/>

論点1:必要な表象を全て記述できるのか？

(1)松原先生：

チェスも最初はとても人間らしさを強く含んだ問題だと思っていた。今でもそういう部分はあると感じているが、一度「**盤面評価**」等の形で記述できてしまうと後は**計算機パワーの問題が主要因**になるようになってしまった。高速化/効率化等まだまだ議論の余地は多くあるが、**計算機パワー**でなんとかなくなってしまいそうな状況。それに対し、東大入試で求められる多様な常識は、それをどう記述すれば良いのかすら分からない段階。

(2)新井先生：

細分化されたことで意識にも登らなくなっている観点やタスクが数多く見逃されているはず。それを効果的に見える化できるように、**適度な粒度・サイズに分割しやすい題材として東大入試を選んだ**。これをオープンな開発基盤として提供する事で**オールジャパン**のアンブレラとして他分野の融合を図りたい。センター入試に5年、東大入試に10年と区切りを付けて遮二無二取り組むことで shake に繋がると思う。

(3)安西先生：

確かにやってみないと分からないことはあるので、これまでになかった新たなアイデアがでてくることもあるとは思う。いずれでも、**これまでも様々な表象を記述する議論であったり、データベースとして格納するといった議論はずーっと昔から数多くやられているが一向に解決する様子が見えない**。ロジックだと記述論理等があるがそれで本当に全て記述できるのか。(経験から学ぶ)フィードバック系だと様々な系をくっ付けることは可能だがそれをどう接続するのが適切なのか。ここまで死屍累々の状況を鑑みると、**オールジャパンで取り組んでも全部が全部ダメでした**で終わってしまったりしないか。

(4)新井先生：

そこまで正面から対峙する必要(全ての表象を記述する必要)があるのかという観点もある。例えば、学生を見ていると「経験から学び、抽象化してそれを次に活かしているとは*思えない*事例」も少なくない。

出典:「人工頭脳プロジェクト(ロボットは東大に入れるか) キックオフシンポジウム

(2011/12/14)、パネルディスカッションメモ: <http://ie.u-ryukyu.ac.jp/tnal/archives/1859>

論点2:情報爆発への対峙の仕方に 「常識」で絞り込む？

安西先生：

初めて対峙する際、例えば初めてある問題について学ぶという時点では丁寧に文章を噛み砕いていくしかない。それこそ自然言語処理のように。ただ慣れてくるにつれてキーを簡単に見つけ出せるようになる。要約みたいなものかもしれない。そこに新たな研究テーマはあるのか。

新井先生：

昨今どこにでも情報爆発は見られてて、それに対する解決策は「気のきいたように処理してあげる」しかないように思う。今の若い世代の人は体感的にそういう問題意識を持っていて、これを解決しない事にはこれ以上の進展が無いだろうという直感がある。身体性が持つある種の感性や感覚が渾然一体となっている常識を解決すること自体が求められている。

安西先生：

研究としてやるとすると、コンテキストチュアルからキーを見つけるなりして、そこから検索に繋げるような感じかもしれないが、それはどう実現するのか。

出典：「人工頭脳プロジェクト(ロボットは東大に入れるか)」キックオフシンポジウム

(2011/12/14)、パネルディスカッションメモ: <http://ie.u-ryukyu.ac.jp/tnal/archives/1859>

ANPI NLP(東日本大震災のためのデータマイニング・自然言語処理)でのお手伝い

- ・ 楽天技研の萩原さんによるツイート
「こういう緊急時にも自然言語処理が役に立てることはたくさんある。情報抽出、データマイニング、情報の信頼性判定、同時翻訳、等々。落ち着いたら自分たちに何ができるかを改めて考えてみるのも良いかも。」
- ・ 言語処理屋さんが連携していく様子(2011.3.14)
<http://togetter.com/li/111529>
- ・ 連携様子その2(2011.3.15~16)
<http://togetter.com/li/112372>

- ・ ANPI NLP:
http://trans-aid.jp/ANPI_NLP/
- ・ メインタスク
 - Twitterなどから得られる安否確認情報を照合、更新することでPerson Finderの情報を充実させること
- ・ 東北関東大震災とソーシャル・メディア (ショートバージョン)
<http://ie.u-ryukyu.ac.jp/tnal/archives/1548>

1
年
後

言語処理学会(2012.3.16)
テーマセッション:「災害時における言語情報処理」(全体討論)
→災害前/直後/復興段階に応じた分析・タスク設計?
言語処理以外との連携?
出典: <http://ie.u-ryukyu.ac.jp/tnal/archives/2042#A5-discuss>

〈具体的にやってる/やりたい〉
事例集

～NAL研～

コンテキスト推定

401教室に関するパネリストの話（実況ツイート）

1. 401は狭かったのでB2は管理が大変だった。多分毎年のこと。 #ieryukyu
2. 他学科や他の大学では自習室なんてない。なのに、規律が守られてないのでせつかくの特権がなくなる。意識向上が必要。 #ieryukyu

↓ 上記の話題提供を受けての関連ツイート抜粋

3. ふらっと、4階に行ったら、本当に学習に使っているか怪しいことが多々。 #ieryukyu
4. 401は正直狭いです。 #ieryukyu
5. 汚い・私物ありすぎ、両方の理由で使ってませんでした。 #ieryukyu
6. 私物が多かったから使ってなかった。 >401 #ieryukyu
【Mac】
7. 2年次の際は使ってなかったな。部屋も狭かったし、すごい汚かったし。 #ieryukyu
8. カバンに財布おきっぱで盗られないってのも平和ですごい事とは思う #ieryukyu
9. 狭いなら、406を利用してよいのでは？ #ieryukyu
10. OSのシーズンの401は狭さもあいまって地獄絵図でした #ieryukyu
11. 狭い部屋に効率良く人を詰めれるように机を配置換えとか...では解決できないレベルの狭さかな。 #ieryukyu
12. 独占は許されないんだけども独占しないとスペースが確保出来ないっていう #ieryukyu

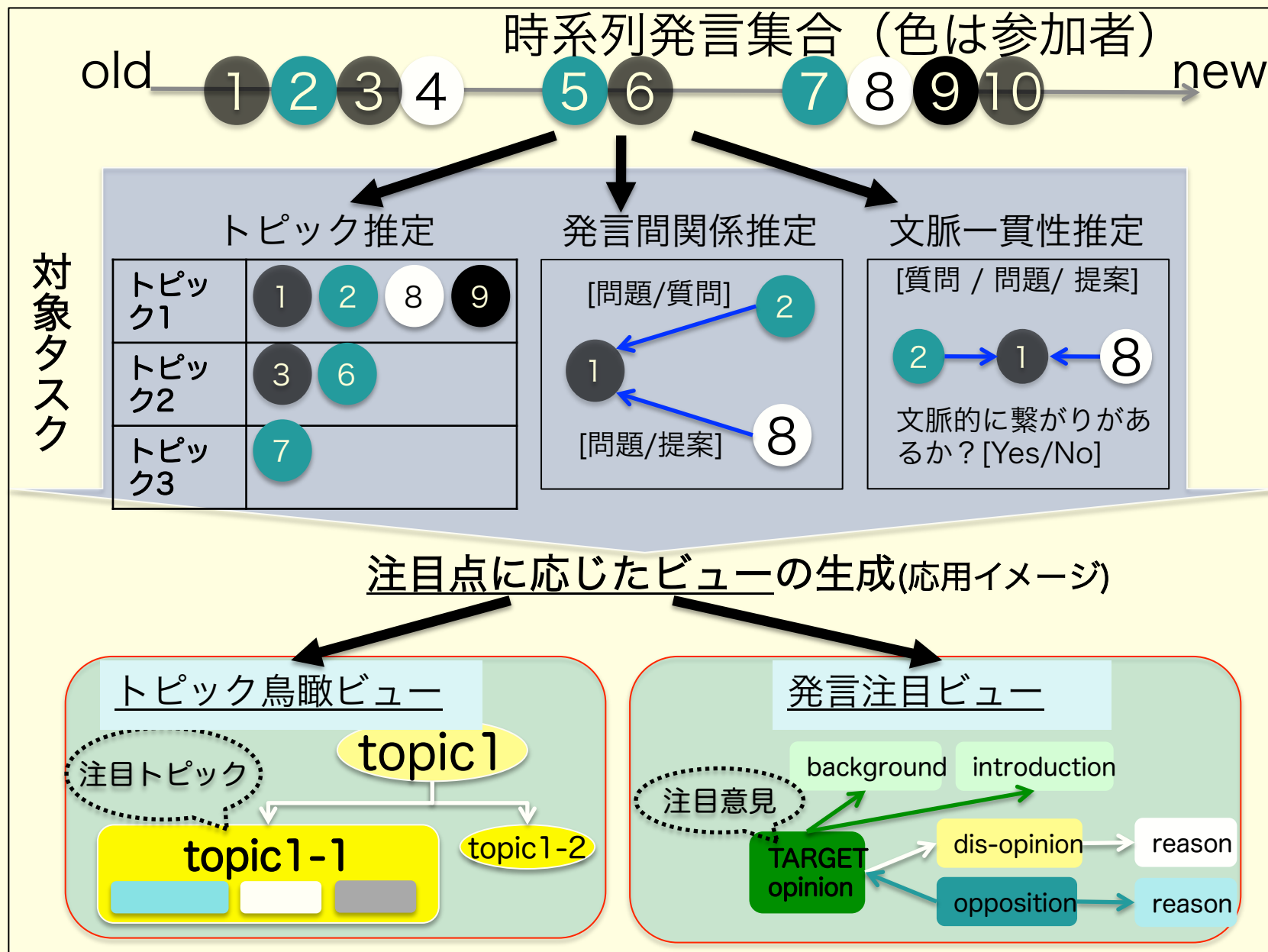
推定して欲しいコンテキスト例

- 「401」は「4階/自習室/私物/汚い/部屋/机」等に関連があること。
- 否定的評価「狭い/汚い」が現れていること。
- 肯定的評価「平和/すごい」が現れていること。
- 一般概念コーパスより「自習室→学校や部屋」への関連する話題であること。
- 「狭さ」を含むツイート群が多く、「部屋」の「狭さ」に関する話題が多く、次いで「部屋」の「汚さ」に関する話題が生じていること。
- 狭さについて「地獄絵図」、「解決できないレベルの狭さ」等の形容度合いが例示されていること。
- 狭さに対する解決策として「406を利用してよいのでは」、「机を配置換え」等が例示されていること。

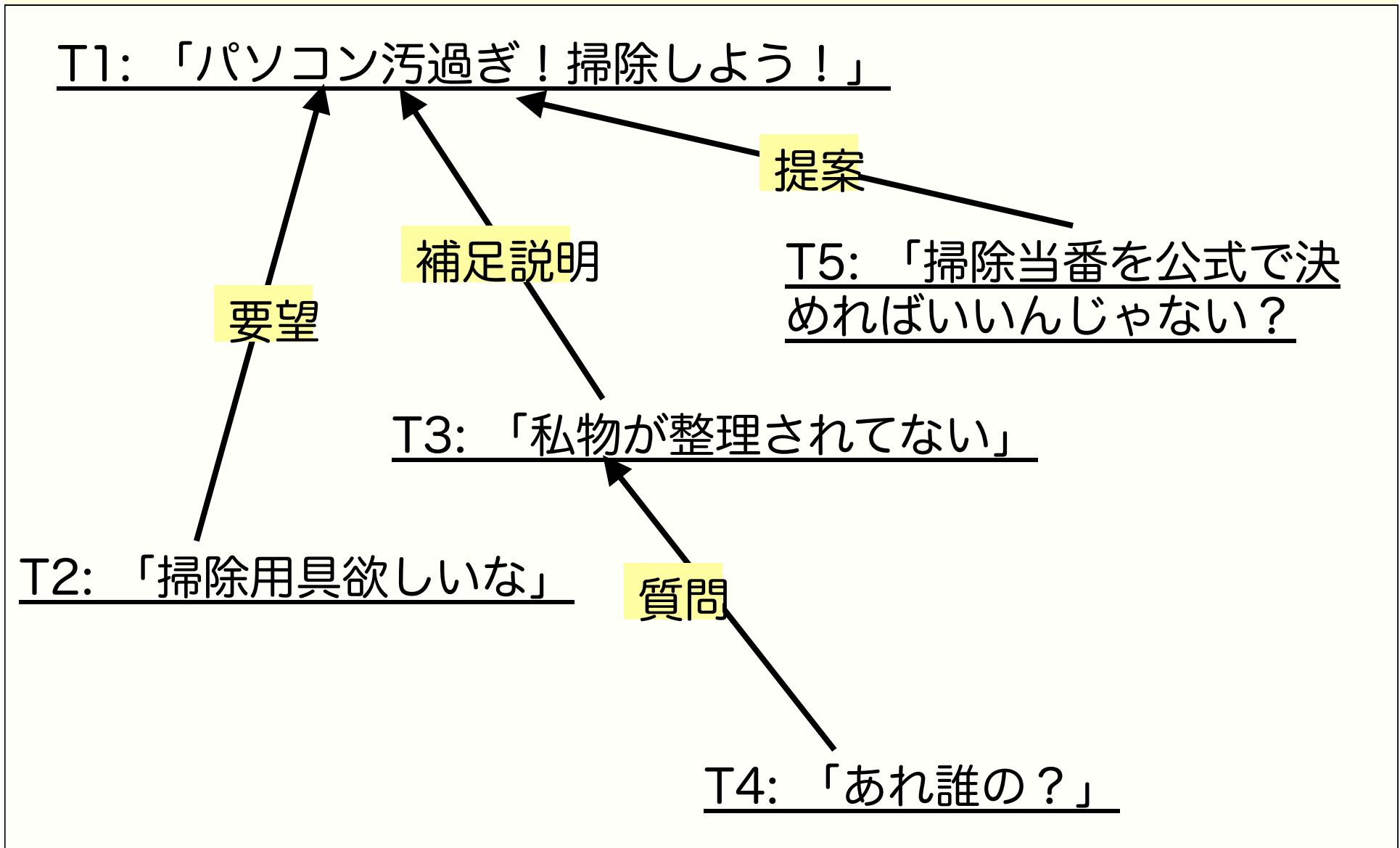
•狭い

- 教室小さい
- 講義登録人数の上限も上げて欲しい
- 20単位縛り

議論構造獲得



発言間関係の例



NAL研テーマ

- ・ (人間的な) 仮想秘書
- ・ 子どものように育てられるシステム

<やりたいこと>

- ・ イベント情報のための情報検索・推薦
- ・ サマリープロジェクト

<技術要素>

- ・ 自然言語処理
- ・ Webインテリジェンス

<やりたいこと>

- ・ 自然言語を介したやり取り (柔軟なUI)
- ・ 常識{の獲得, を利用した応用}

<技術要素>

- ・ ALife/AI/自己組織化/創発/身体知
- ・ 仮想身体? 実世界へのリンク?

背景や動機

- 自前のweb, mixi, BBS, Mail, iCal, txt, PowerPoint, iTunes, iPhoto, FreeMind,,
 - 様々な情報が電子化されているのに、自分がコンピュータに合わせて作業をしないといけないのは何故だ？
 - 自分の好み・嗜好を汲み取って、前もってチェックした情報を整理して提示してくれないか？
- Web上の大量のコンテンツ群
 - Web検索で検索結果リストから探しているページかどうかをチェックするのも面倒！
 - 「9月に沖縄に旅行へ行きたい。2泊3日で、ホテルは恩納村。旅費は空路と宿泊費で7万円ぐらいで、海に近い方が良い」

関連既存サービス

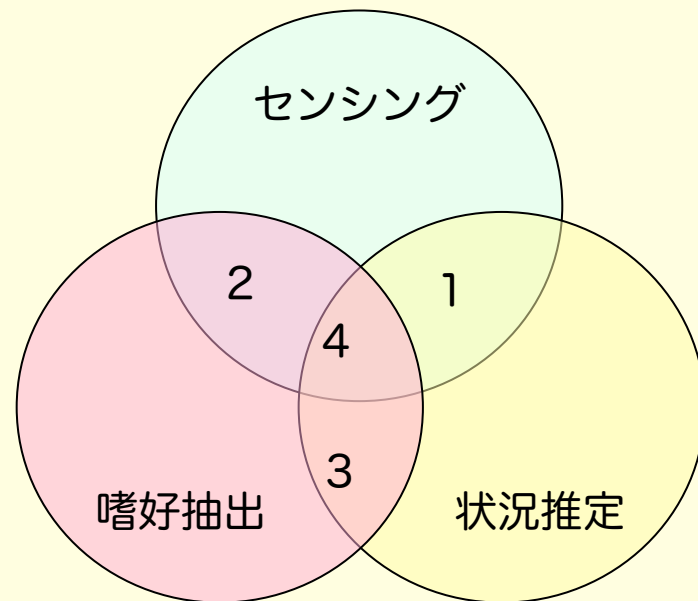
- 情報(商品)推薦: Amazon等
 - 利点
 - 興味のあるような商品を推薦してくれる
 - 欠点
 - 類似商品ばかり推薦され過ぎ
 - 誰かが評価してくれないと推薦候補にならない
 - SPAM等による作為的な操作
- 企業が儲けるために商品推薦するサービス
 - 顧客から集めた情報を企業のため（企業主体）に使っている
 - 顧客自身のためには使っていない

提案:

ユーザ主体の高度なサービス（仮想秘書）の実現

より高度な情報検索・収集・整理・推薦・提案…

どの程度の質・量の情報を用意出来れば、どのようなサービスを実現可能か



関連キーワード

CGM, ライフログ, 予測, 推定, セマンティックウェブ, 情報統合, 自然言語解析, 対話, 戦略生成, ,

ゴールのイメージ例

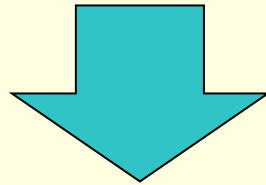
- ・ より高度な情報検索
 - 結果を「ページ」ではなく「コンテンツやサマリ」で返す。
 - 自然言語で質問できる。
 - 必要ならば推論を交えた回答を用意する。
 - ・ 例: サーバ班議事録

- ・ 少し具体的なゴール
 - AutoReferer
 - GhostWriter
 - スケーラブル・サマリー
 - FormatAdvisor
 - ...

典型的なレコメンドシステム

【背景】 膨大なアイテム集合の中から、できるだけ労せず、希望にマッチしたアイテムを探し出したい。

【解決】 購買履歴等の**評価情報**や**顧客情報**を基に、**アイテム**に関する好みや傾向を分析し、関連度の高いアイテムを推薦。



イベント情報に求められるレコメンドシステム

【条件1】 イベント情報に対する**評価を得られない状況**で機能するモデルであること。

【条件2】 観光客等のように、必ずしもその地域の**ローカル情報**に**詳しく無い状況**でも機能するモデルであること。

【解決】 ユーザセグメントとイベントセグメントを生成し、それらを紐付ける事で、**ユーザ視点で絞り込みやすくする**。