

ソフトウェア開発の議論における タグ付けを利用した情報統合の提案

学籍番号：1421030 氏名：石川 俊明

指導教員：鷹野孝典 准教授

研究背景

- ソフトウェア開発では、設計や実装に関する情報共有をする際、参加者各自でメモを取ったり、議事録を残すことで必要な情報を記録している。
- 様々な場所で議論をしたり、議論のメンバーが変わる事で、同じ組織内で似たような話を何度もする事になったり、十分に共有出来たか確認出来ないという問題がある。



関連研究

1. 『履歴再生機能を備えたオンラインホワイトボード・チャット連携システム』
(2010/05/14, 情報処理学会 教育学習支援情報システム (CLE) 研究会)
2. 『ホワイトボードを用いた議論の構造化に基づく議論想起支援』
(2011/03/02, 情報処理学会 全国大会講演論文集)
3. 『簡易書式によるホワイトボード動画を対象としたインデックス抽出』
(2011/03/02, 情報処理学会 全国大会講演論文集)
4. 『書き込みの時間軸表示によるホワイトボードログの振り返り支援システムの実装』 (2012/09/14, 情報処理学会 電子化知的財産・社会基盤研究会)

議論の関連研究

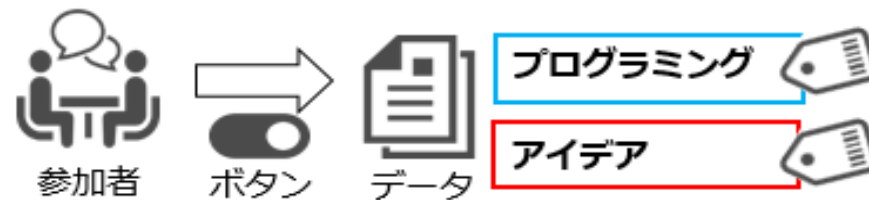
1. 『知識活動支援システムによる会議コンテンツ間の関連性の獲得とその応用』
(2010,情報処理学会 全国大会講演論文集)
2. 『議論マイニングによる議論掲示板利用者の能力推定』
(2017,情報処理学会 全国大会講演論文集)
3. 『参加者の議論能力と役割を考慮したオンライン議論の分析』
(2013,人工知能学会 全国大会論文集)
4. 『係り受け構造を用いたWeb議論掲示板における投稿への自動分類』
(2016,電子情報通信学会 技術研究報告)

研究動機

- ソフトウェア開発では、開発や設計に関する情報を共有するために様々な手段を利用して、意見交換を何度もする。
- その度に「なぜこの技術なんだっけ？」「この人に実装を任せれば早く終わったね」といった振り返りをする事がある。
- 手段に依存せず、過去に取り上げた問題解決の手法やアイデアを確認する事で、現在の問題を素早く解決したい。

本研究の目的とアプローチ

- 目的：議論の手段に依存せず、議論の情報を集約可能な手段の検討と開発
- アプローチ：様々な場所で利用可能な「物理ボタン」を利用し、議論にタグを付けて議論を記録する.



- 現在と過去の議論をタグで紐付け、新しい議論で情報として利用する.
- 議論の手段に依存せず、議論内容の記録が可能にする（例：テキストチャットで行なった開発議論と口頭での開発議論の紐付け）

議論をタグに置き換える例

- ゼミの場合

- 「研究のアイデア」
- 「実装のアイデア」
- 「情報共有のあり方」

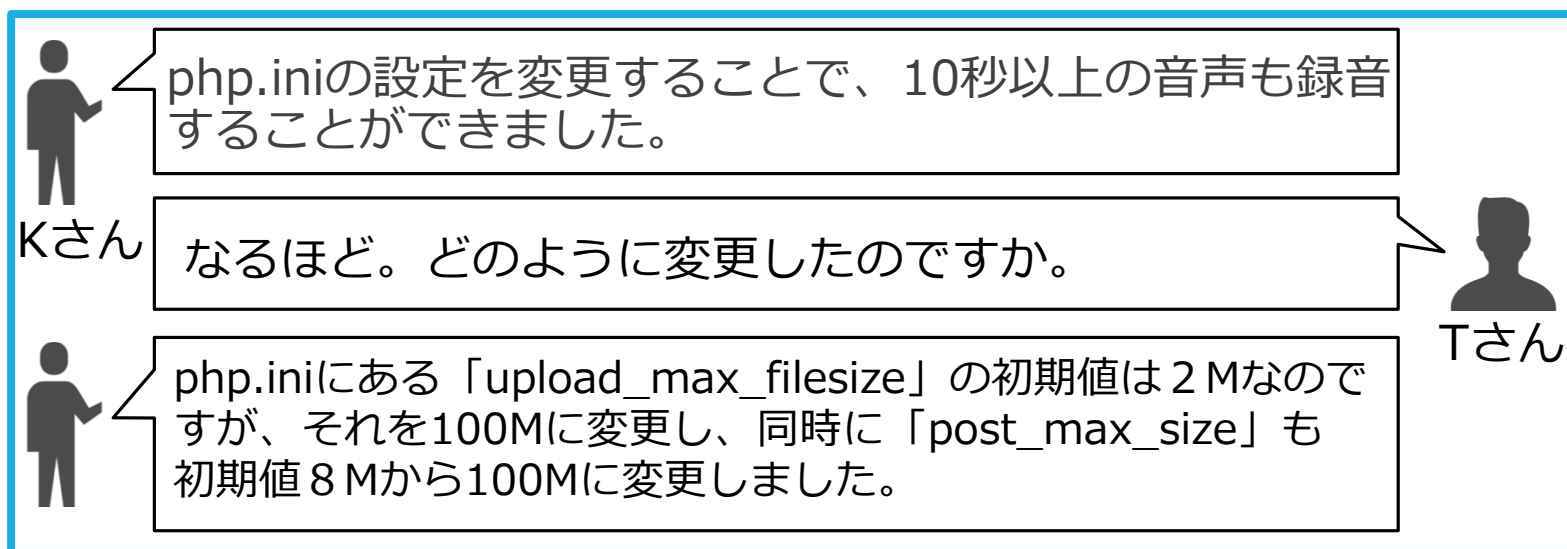
→タグとして考えると「研究」「実装」「アイデア」「アドバイス」などが挙げられる。

- 勉強会の場合

- 「PHPの使い方」
- 「JavaScriptの使い方」
- 「Slackの使い方」
- 「GitHubの使い方」

→タグとして考えると「プログラミング」「Webサービス」が挙げられる

実際の議論（会話）例



- 60個の発言からタグを付けられそうな会話群は3件（19個の発言）。
- 上記例の場合、「**PHP**」や「**サーバー設定**」といったタグで記録することで、振り返り時に利用出来るという利点や、「**サーバー設定**」に関して知識のある人物の発見が出来るのではないか。

従来方式と比較

【メモで記録する場合】

- アイデアそのものや決定した方式そのものが記録される

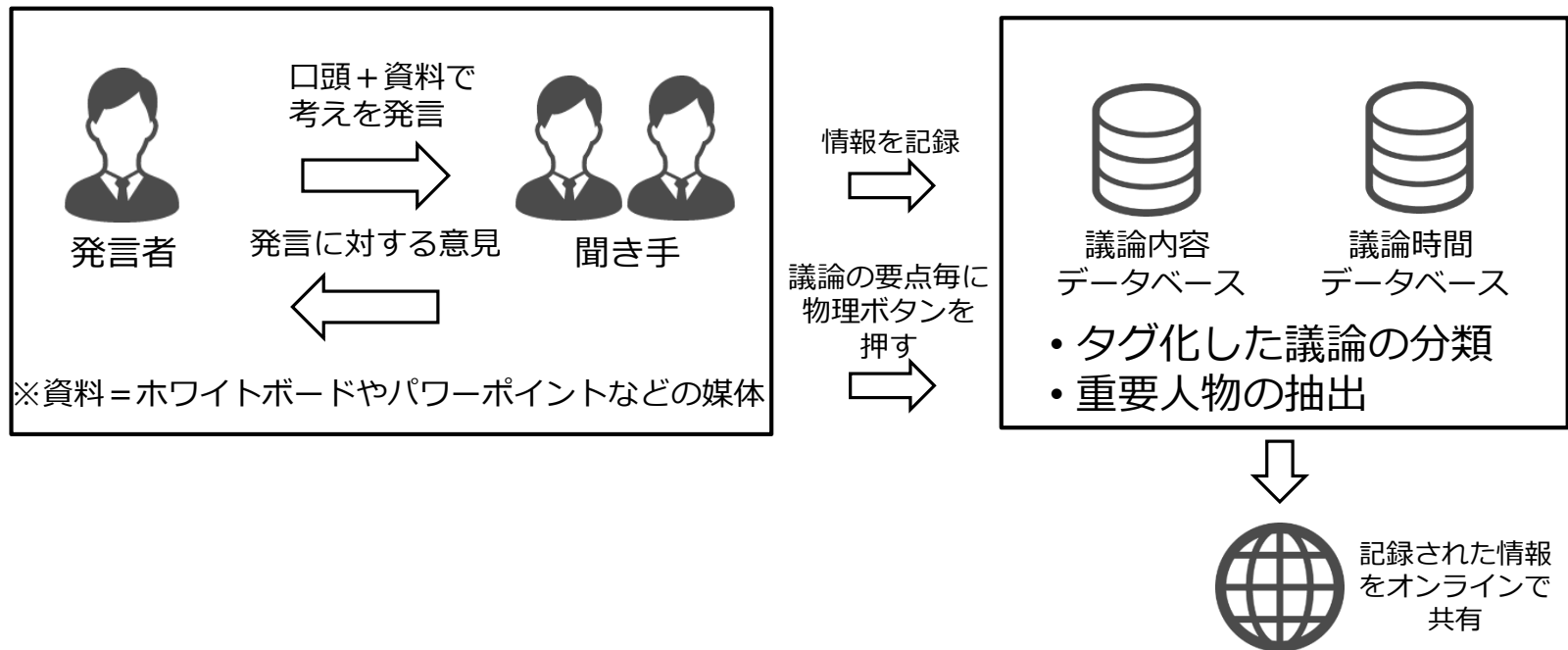
【本提案方式の場合】

- アイデアを提案した人や実装方式に関する一連の議論が記録される

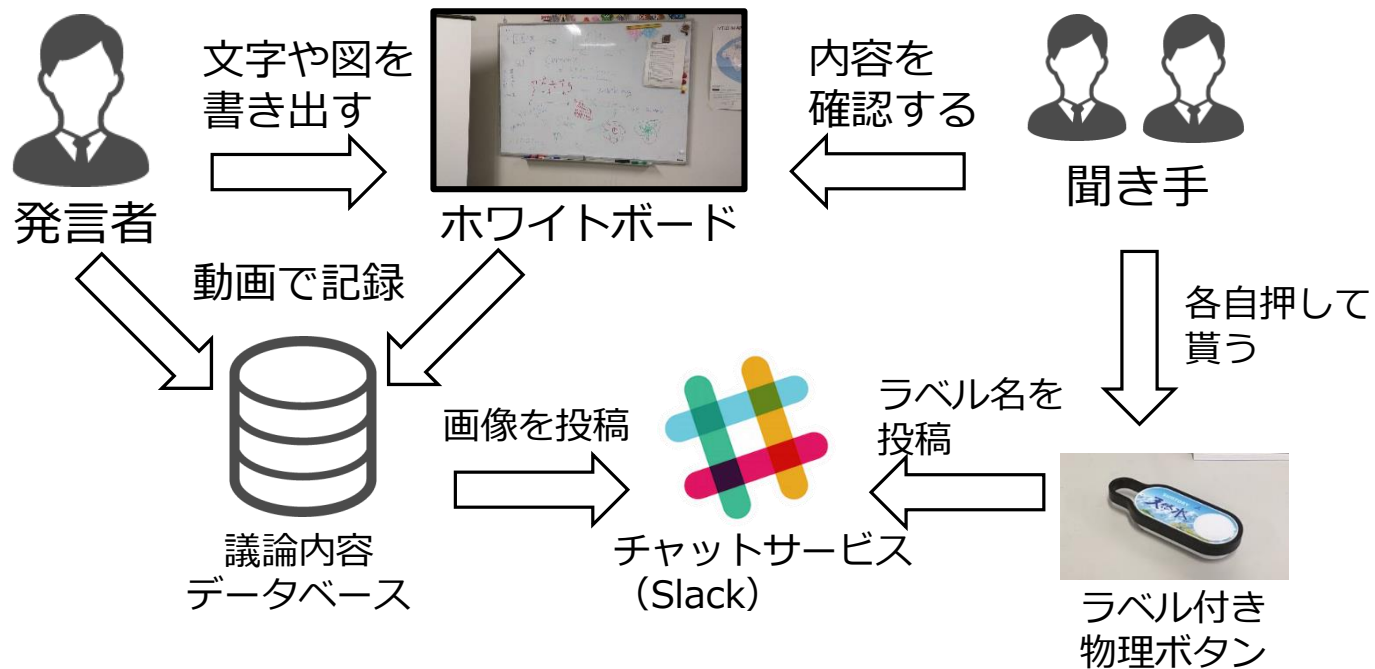
→ 本提案方式によって「アイデアを提案した人物」や「実装方式の利点」が記録される。

提案システム図

議論の場



システムの流れ図



- オンラインでの共有として、研究室で利用しているチャットサービスを利用する

実装状況

- ラベル付けされた物理ボタンが押されたときに， Slackへホワイトボードの画像を投稿するプログラム
- 物理ボタンを押した時にその時間をデータベースに保存するプログラム
- 物理ボタンの代替となる記録用ページの作成



物理ボタン

タグ一覧

実装

設計

プログラミング

未解決

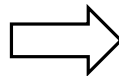
解決

保存

記録用ページの作成

Slackで実装する時の問題

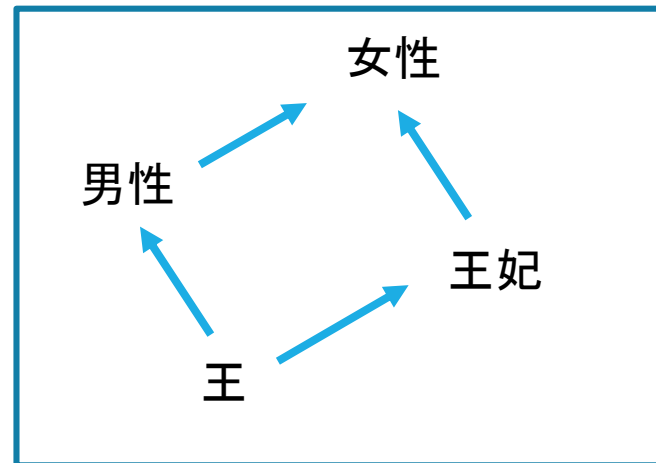
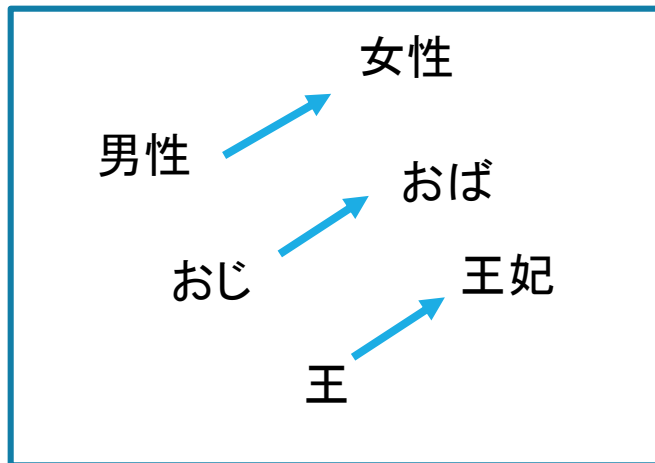
- 物理ボタンにチャンネルの紐付けが出来ない。
- APIで全てのメッセージを受け取る仕組みがあるものの、流れてくるどのチャットが物理ボタンと紐付けているのか分かりづらい可能性がある（ここは実装方法次第で解決出来るかも）



```
[{'source_team': 'T0J8LKSTV', 'text': 'test', 'ts':  
u'1508827729.000156', 'user': 'U106UP90U', 'team':  
u'T0J8LKSTV', 'type': 'message', 'channel': 'C7BRCTVUP'}]
```

Doc2Vecで関連度抽出

- (前提技術) Word2Vecについて
 - 単語をベクトル (数値配列) で表現する技術のこと
 - 距離が近い単語は意味が似ていて、遠いと似ていない

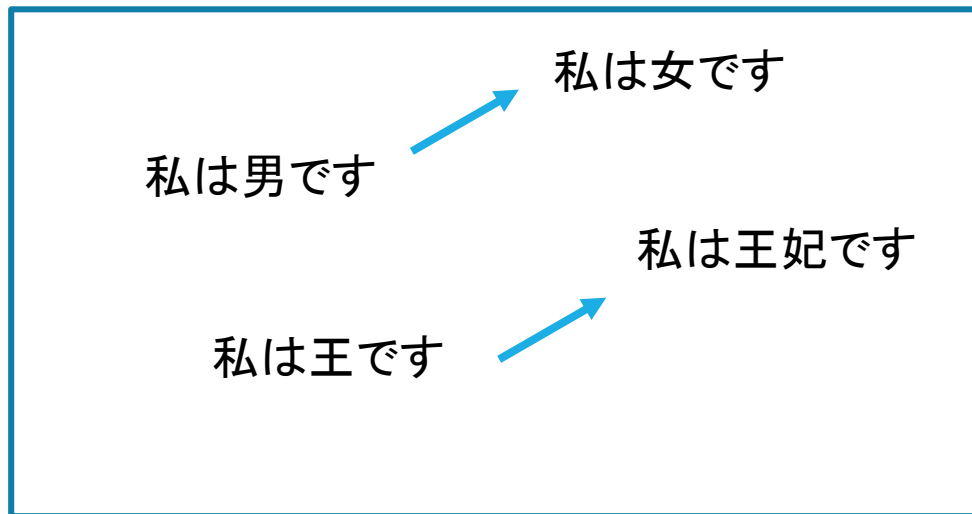


- 例1) $\text{王} - \text{男性} = \text{権力のある人}$
 $\text{権力のある人} + \text{女性} = \text{王妃}$
例2) $(\text{良い} + \text{最高}) / 2 = \text{より良い}$

Doc2Vecで関連度抽出

- **Doc2Vec**について

- Word2Vecの技術と同様に、文書をベクトル（パラグラフベクトル）で表現する技術
- 可変長の文章や段落に対応するので文章と文章の関連度の抽出が出来る



- 文章の要約やニュース記事の類似度抽出などといった活用法がある

検証

- 以下の文章を分かち書きした上で日常的な文章10個を含めて関連度を検証
 - 「php.iniの設定を変更することで、10秒以上の音声も録音することができました。」
 - 「php.iniファイルの変更が必要です」
- 一番関連している文章として抽出する事が出来たものの、値としては低い (0.29…).
- 何度か試行すると別の結果が出てくる事もあり、短い文章だと有効性があまり示せず、チャットの場合は短い文章で会話が続く事もあるのでDoc2Vecだと上手く抽出出来ないかもしれない事が分かりました.

今後の課題

- 現在と過去を結びつける良いタグが見つかるとは限らない.
- 音声からのラベル付けは音声認識の研究になってしまう.
- タグ毎の投稿数から「重要度」を取るのか, タグを付けた人から「重要度」を取るのかが定まっていない.
- 従来の手法の設定と提案方式の比較方法について
 - 個人が記録するメモなどでは議論で決まった事のみがメモされるのが殆ど.
 - 提案方式によってその決まるまでに至ったプロセスを示す事が出来る(?)