言語グリッド: サービス指向集合知による多言語サービス基盤

村上　陽平

（独）情報通信機構　言語グリッドプロジェクト

2010年10月7日　CEATEC
Webサービス、クラウドの先へ：
サービスコンピューティング研究が拓く世界

翻訳サービスを使ってみる

- 日韓翻訳
  - 「今日は君が掃除当番だよ.」
  - 「오늘은 너가 청소 당번이야.」
  - 「今日はお前が掃除当番だ.」
- 日中翻訳
  - 「今日は君が掃除当番だよ.」
  - 「今天你是扫除值日哟.」
  - 「今日あなたは取り除いて当番をするのです.」
- 複数の資源を組み合わせてカスタマイズできない.
  - 辞書を登録できない
  - 翻訳例文を登録できない
  - 現場では翻訳品質を受け入れざる得ない → 使わない

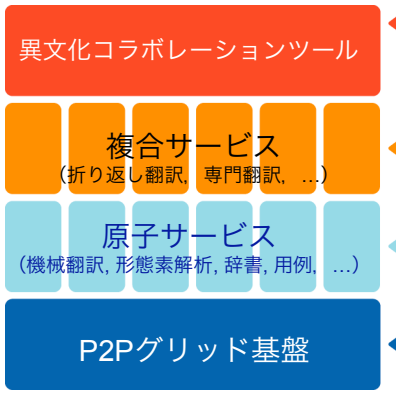
言語グリッド

目次

- 言語グリッドの基盤
- 言語グリッドの運用
- 言語グリッドを通したサービスコンピューティングの課題

言語グリッドの基盤

言語グリッドのサービス階層



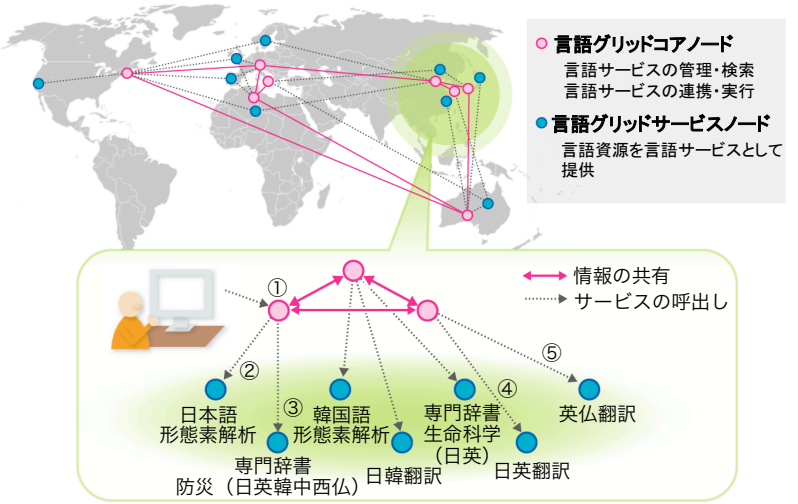
異文化コラボレーションを支援
言語サービスを利用し，多言語環境でコミュニケーションを支援．NICTでは言語グリッドコラボレーションツールを提供．

言語資源の連携
複数の言語資源（Webサービス）を，ワークフローを用いて連携．

言語資源の利用
言語資源を標準インタフェースを持つWebサービスとして提供．

計算資源の管理
インターネット上の複数のサーバを連携させ，言語資源，計算資源を世界中からアクセス可能に

P2Pグリッド基盤

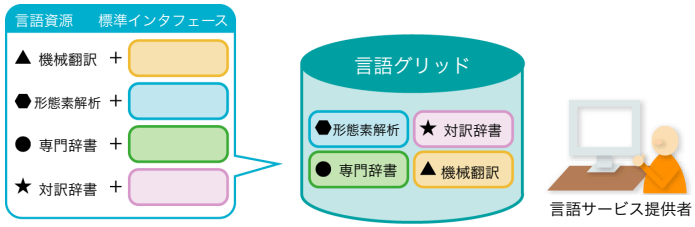


イメージ図

原子サービス

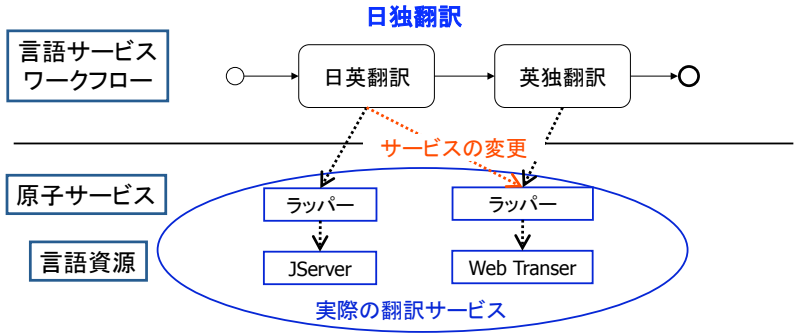
Webサービス

- 各言語資源の種類ごとに定義されたインタフェースでWebサービスとしてラッピング
- ラッピングの負荷を軽減するために共通する処理をラッピングライブラリとして提供
- 言語サービスのインタフェースの標準化が必要



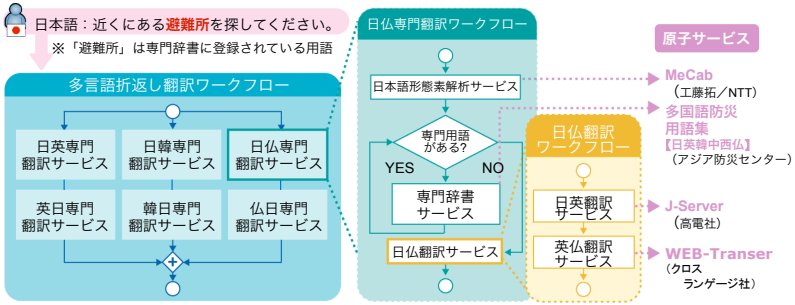
複合サービス

- 新しい言語サービスをワークフロー記述言語WS-BPEL（業界標準）で開発.
- ワークフローは言語資源のインタフェースと制御構造から構成
- 利用する原子サービスを実行時にバインディング(SOAPヘッダで指定)
- 複数サービスの併用が可能



複合サービス

異文化コラボレーションツールのための言語サービス
～ 防災時の英語・韓国語・仏語への日本語の翻訳 ～



翻訳結果

英　語：Please look for the shelter in the neighborhood.
韓国語：가까이 있는 피난소를 찾아 주십시오.
仏　語：Veuillez rechercher le refuge dans le voisinage.

折り返し翻訳結果

どうぞ、付近の避難所を捜しなさい。
近くある避難所を探して下さい。
どうぞ、付近の避難所を検索しなさい。

異文化コラボレーションツール Language Grid Toolbox

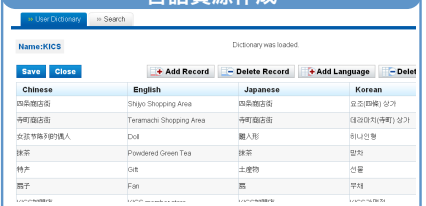
テキスト翻訳

・折り返し翻訳機能により、翻訳結果の内容確認ができる

多言語掲示板

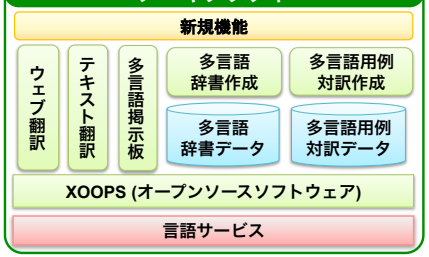
・多言語に翻訳される掲示板
・翻訳結果を人により修正し、翻訳の品質を向上できる

言語資源作成



・機械翻訳と連携し利用するためのコミュニティ辞書や対訳集の作成ができるツール

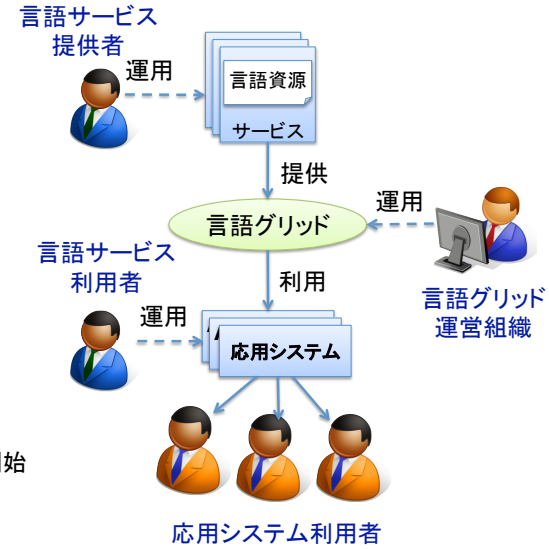
アーキテクチャ



言語グリッドの運営

中央集権型運営モデル

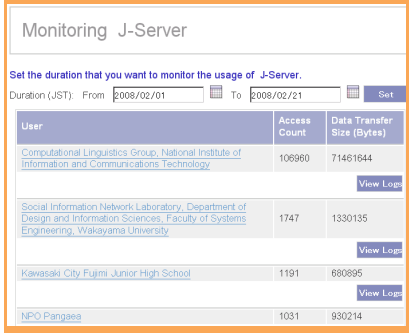
- ステークホルダー
  - 言語グリッド運営組織
  - 言語サービス提供者
  - 言語サービス利用者
  - 応用システム利用者
- 言語グリッド運営組織を中心として言語サービス提供者および言語サービス利用者が覚書を締結するモデル
  - 非営利目的に限定
  - 2007年12月より運営開始



言語グリッドサービスマネージャ

利用者および言語グリッド上の言語資源，言語サービスを管理するためのWebベースの管理ツール

言語資源の利用状況のモニタリング



言語資源のアクセス制約設定

多言語サービス基盤の構築

参加組織（18カ国131組織が参加）

【大学】
大阪大学，関西大学，関西学院大学，京都大学，東北大学，長岡技術科学大学，名古屋大学，はこだて未来大学，北海道大学，立命館大学，早稲田大学，和歌山大学
カトリック大学（韓国），韓国国民大学（韓国），上海交通大学（中国），インドネシア大学，シュツットガルト大学（ドイツ），清華大学（中国），プリンストン大学（アメリカ），ケベック大学（カナダ），コペンハーゲン大学（デンマーク）など

【研究機関】
DFKI（ドイツ），CNR（イタリア），中国科学院，NECTEC (タイ)，国立情報学研究所，NTT研究所など

【NPO/NGO】
愛知県国際交流協会，アジア防災センター，NPOパンゲア，NPO多文化共生センターきょうと，川崎市総合教育委員会，川崎市立富士見中学校，多言語防災情報研究コンソーシアムなど

【企業】（社会貢献または言語資源の提供）
Google inc.，東芝，沖電気など

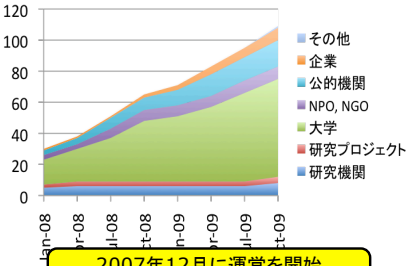
言語サービス（90以上のサービスを共有）

【機械翻訳】
Google Translate（51言語），J-Server（日英韓中），WEB-Transer（日中韓英仏独伊西葡），東芝（英中），沖電気（日英），Parsit (英->泰)

【対訳辞書】
EDR, ライフサイエンス辞書，学術辞書，防災用語集など

【用例集】
医療用例対訳集，教育用例対訳集など

【形態素解析】
日中韓英仏独伊西蘭葡露

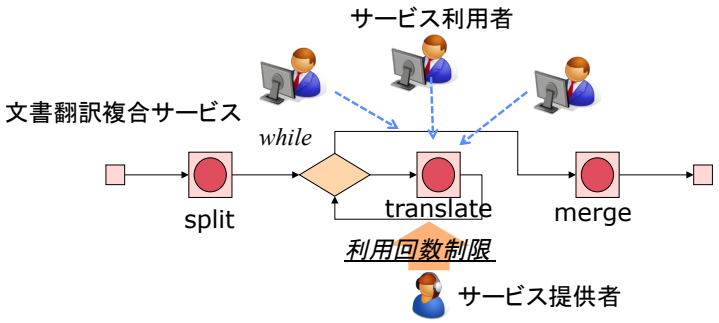


2007年12月に運営を開始
3年間で順調に成長している

言語グリッドの運用を通した サービスコンピューティングの課題

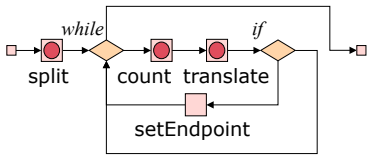
サービス実行時の環境適応

- サービスは環境によって振る舞いが変化
  - サービスの割り当て，QoSの変化
- e.g. 利用回数の制限への対処
  - 言語グリッドユーザグループに対して実行回数制限を設定
  - あるユーザで制限を超過すると，全ての実行インスタンスで実行失敗



問題：複合サービスの改変不可

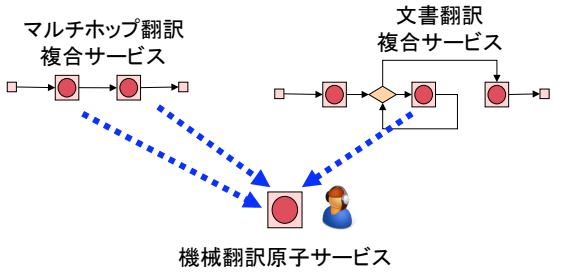
- 複合サービスに処理を加えて対策
  - 全インスタンスでの合計利用回数の記録
  - 利用可能回数を超える前にサービス切り替え



- オープンな環境ではユーザ・複合サービス設計者が別
  - ユーザは複合サービス改変の権利がない（複合サービスの知的財産権保護を考慮）
  - 設計段階では，どのような処理が必要が不明

問題：改変コスト

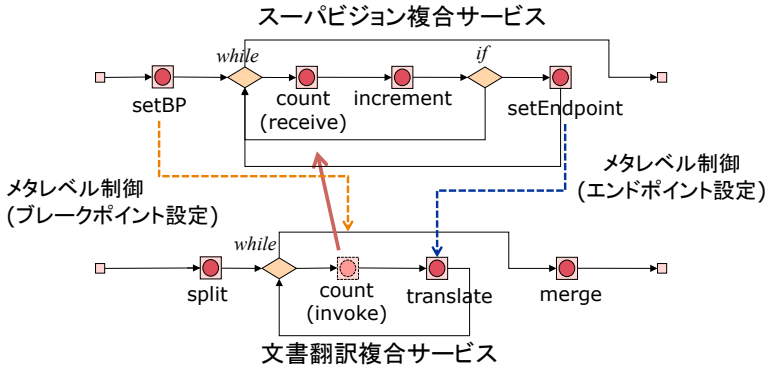
- 原子サービスはさまざまな複合サービスから利用されうる
  - 翻訳サービスの利用：マルチホップ翻訳，文書翻訳，辞書連携翻訳



- 改変可能な場合でも，関連する複合サービスの全てについて，必要な処理を加えるのは大変

スーパビジョン複合サービス

- 制御を複合サービスとして定義
  - 制御対象から分離・再利用可能
  - 言語グリッド運営組織が記述



複合サービスのQoS

- 言語グリッドのような大規模サービス連携基盤の問題
  - 言語サービス提供者と言語サービス利用者は物理的に分散
  - サービス実体と言語サービス利用者の物理的距離が遠ければサービスレスポンスの遅延(QoS)が大きい
    (物理的な距離がボトルネック)
  - サービス利用者は複合サービスを利用する場合、各原子サービスが分散すれば、レスポンス遅延(QoS)はさらに大きい

言語資源提供者の物理的な場所

問題:複合サービスの応答速度

- サービス呼び出し時間(ms)比較(JGN2plus上の言語グリッド)

	J-Server翻訳サービス (タイからの呼び出し)		辞書連携複合サービス (J-Server+Mecab+京都観光辞書) (タイからの呼び出し)	
	サービス実体(タイ)	サービス実体(日本)	サービス実体(タイ)	サービス実体(日本)
1	200	578	285	1474
2	126	413	290	1429
3	179	421	310	1437
4	119	408	293	1446
5	184	405	250	1438
6	184	410	326	1452
7	129	417	264	1444
8	191	411	411	1486
9	186	411	253	1440
10	188	409	293	1434
平均	168.6	428.3	297.5	1448

遅延2.54倍　　遅延4.87倍

問題:ライセンスによる制約

- 大規模サービス連携基盤のサービス配置問題を解決するには、既存の手法の適応が困難
  - なぜサービス実体を全てのサービス実体ホストに複製できないか
    - サービス実体のライセンスなどによって、サービス実体の配置に制限がある場合が多い
  - なぜCDNのキャッシング技術が活用できないか
    - データと比べると、サービス実行結果のヒット率が低い
    - 特に複合サービスの場合は、ヒット率がさらに低い

ユーザのリクエストに応じてサービス実体の配置を動的に変更し，ユーザにとってQoS(反応速度)の高い多言語サービスを提供

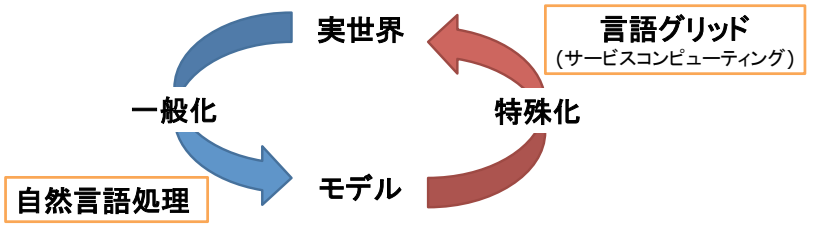
サービスのグループ特性に基づく動的再配置

- 連携させるサービスの配置を変更
  - サービス利用者近くのコアノード近傍のサービスノードにサービスを集約
  - 利用履歴に基づき，利用回数の多いサービス利用者を優先
- 原子サービスのグループ特性を考慮した動的再配置
  - 複合サービスを構成する原子サービスをグループとして移動
  - 独英翻訳サービス、英泰翻訳サービス(ドイツに移動)
  - ドイツ語形態素解析サービス、独泰専門辞書サービス(ドイツに移動)

言語グリッドへの参加募集！

- サービスコンピューティングを始めたい方
  - インタフェースが標準化された90以上のサービス
  - 企業の方も研究目的で利用が可能に！
- 異文化コラボレーションを支援したい方
  - 多言語サービスが無償で利用可能
  - ドメインに特化した言語資源も利用可能
- http://langrid.org/operation
- operation@langrid.org

まとめ



- 異文化コラボレーションの現場には汎用的な言語資源や言語処理ツールではなく，カスタマイズされた言語サービスが必要
  - 膨大な言語資源の所有から必要な言語サービスの利用と結合を可能にする言語グリッドの構築
  - 利用組織を拡大するために連邦制運営モデルを構築
- 言語グリッドの運用を通したサービスコンピューティングの課題
  - 複合サービス実行時の環境変化に適応するためのサービススーパビジョン
  - 複合サービス実行時の応答速度を改善するためのサービスの再配置