

Linked Dataを用いた クラウドサービス連携アーキテクチャ

M2011MM041

小島 弘誉

+ 概要



- 概要
- 本研究の概要
- 前回の流れと今回の狙い
- 論文の概要
- Scenario and Requirements
- Distributed Organizational Environment
- RDF Update Format
 - ❖ A. SPQRQL 1.1 Update
 - ❖ B. Tails Changeset
 - ❖ C. The Graph Update Ontology
 - ❖ D. The Guranteed RDF Update Format
 - ❖ E. Sesame RDF Transactoins
 - ❖ F. Other formats
- Network Usage
- Conclusion
- まとめと考察
- 今後の課題

+ 研究概要



■ 背景

- ❖ クラウドの普及
- ❖ クラウドサービス連携の必要性

■ 問題: クラウドがサービスデータ連携方法が未定義

■ 研究課題

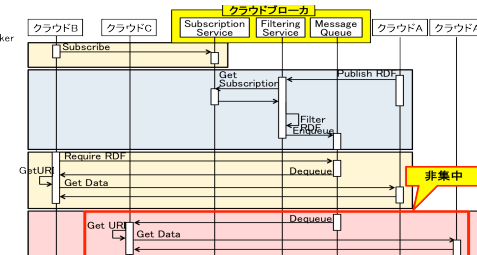
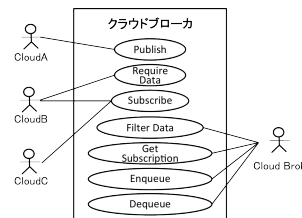
- ❖ 型の整合 (構造の整合, 意味の整合)
- ❖ 振舞いの整合
- ❖ スケーラビリティ: 多量のデータ処理
- ❖ ネットワーク使用量: 様々なサイズのデータ

+ 前回の流れと今回の狙い

■ 前回の流れ

- ❖ 研究の着眼点変更
- ❖ 3つの技術のユースケース図, シーケンス図を記述
- ❖ 振舞い, スケーラビリティ, データサイズでの比較

⇒ Publish/Subscribe with Linked Dataのメリットを明確化



■ 今回の狙い: データ連携のユースケースを学ぶ

- ❖ 読んだ論文:
Approach for real-time integration of Semantic Web data in distributed enterprise systems

+ 論文の概要



■ 背景

- ❖ Web2.0ツールの普及(Wiki, Blog, Microblog)

 - ⇒ 様々なツールで多くのコンテンツを生成

- ❖ セマンティックWeb技術がデータ連携可能性を提供

 - ⇒ 2つの課題が残っている

■ 課題

- ❖ ソーシャルデータのリアルタイム連携

- ❖ 情報の多重定義の制限

■ 提案: 異なるプラットフォーム間でのコンテンツ連携

- ❖ 議論: RDF update format

+ Scenario and Requirements



コンテンツ公開プラットフォームを越えたリアルタイム連携は改善されたコラボレーションと情報の共有を可能にさせる

■ シナリオ

❖ Semantic Webの専門家グループ

✓ 同じ組織に属する

✓ 異なる区分で働く

⇒ 関連のある同僚のブログフォローが必要

❖ RSSの利用

✓ 定期的な情報ソースの問い合わせ可

⇒ アクセス制限されたネットワークでの効果的なデータ連携が許されない

+ Scenario and Requirements



コンテンツ公開プラットフォームを越えたリアルタイム連携は改善されたコラボレーションと情報の共有を可能にさせる

■ シナリオ

❖ Semantic Webの専門家グループ

✓ 同じ組織に属する

✓ 異なる区分で働く

⇒ 関連のある同僚のブログフォローが必要

❖ RSSの利用

✓ 定期的な情報ソースの問い合わせ可

⇒ アクセス制限されたネットワークでの効果的なデータ連携が許されない

+ Scenario and Requirements



■ 具体例

- ❖ SIOC(Semantically-Interlinked Online Communities)
- ❖ Initial annotation of Bob's blog

```
<http://exorople.org/Bob/2010/04/01/social-software> a sioc:Post;  
sioc:topic <http://example.org/category/social-software>;  
sioc:topic <http://example.org/category/semantic-web>.
```

- ❖ Updated annotations of Bob's post

```
<http://exorople.org/Bob/2010/04/01/social-software> a sioc:Post;  
sioc:topic <http://example.org/category/social-software>.
```

■ 要求

- ❖ 良い粒度でのRDF操作
 - ✓ データのCRUD操作
 - ◇ Document level
 - ◇ Statement level

+ Distributed Organizational Environment

■ 分散モデルの現状

- ❖ 現在: 様々なコラボレーションツールを利用(eg. wiki, blog)
- ❖ 課題: 重要なデータは効果的に管理されていない
- ❖ 解決策: RDFの利用
 - ✓ microblog, collaborative informationのキャプチャ
 - ✓ 組織を越えてのRDFコンテンツの連携

■ コミュニケーションモデルの分類

- ❖ Pull: クライアントから呼び出し, サーバが応答
 - ✓ Polling: クライアントがサーバのステータス繰り返しサンプリング
 - ◇ リソースの不効率な利用
 - ◇ 重要なレスポンスが遅れない
 - ◇ スケーラビリティのために非同期配信を求められる
 - ✓ Long Polling: レスポンスが得られるまでリクエストを保持
- ❖ Push: イベントが起こったときにサーバからクライアントへ通知

+ RDF Update Format

RDF dataを用いたreal-time integrationの同期の要求

- ❖ Push interaction model
- ❖ 簡潔なRDF change description

■ SPQRQL 1.1 Update

```
DELETE DATA FROM<http://example.org/.../soc-soft.rdf {  
  <http://example.org/Bob/2010/04/01/social—software> sioc:topic  
  <http://example.org/category/semantic-wcb/> } }
```

■ Tails Changeset

```
<http://example.cora/cs#change> cs:removal _:bl;  
a cs:ChangeSet.  
_:bl rdf:object <http://example.org/category/semantic—web/>;  
  rdf:prcdicate <http://rdfs.org/sioc/ns#topic>;  
  rdf:subject <http://example,org/Bob/2010/04/01/social—software/>;  
a rdf:Statement.
```

- The RDF Update Format
- SesameGraph Upate Ontology
- The Guranteed RDF Transactoins
- Other formats

+ RDF Update Format



Format	Advantage	Disadvantage
SPQRQL Update	<ul style="list-style-type: none">+ W3C standard+ enables CRUD+ compact+ flexible	<ul style="list-style-type: none">- no metadata about the change
Tails Changeset	<ul style="list-style-type: none">+ change metadata	<ul style="list-style-type: none">- requires reifications- Inefficient for large number of triples- a changeset focused on one resource- no dynamic updates
GUO (The Graph Update Ontology)	<ul style="list-style-type: none">+ lightweight+ change metadata	<ul style="list-style-type: none">- focused on one resource- inefficient for large number of triples
GRUF (The Guaranteed RDF Update Format)	<ul style="list-style-type: none">+ very compact+ simplicity	<ul style="list-style-type: none">- no software implementations- no metadata about the change
Sesame RDF transaction	<ul style="list-style-type: none">+ executed as a transaction+ named graph	<ul style="list-style-type: none">- no metadata about the change- not standardized- not documented

+ Network Usage

- ネットワーク利用量はコンテンツの同期の効率性に影響
- 比較概要
 - ❖ 比較: ネットワーク利用量 (実際の利用量の統計に基づく)
 - ❖ User(post: 1027, Comment: 1326, weekly)

	Net Work Usage [kb] with 100 subscribers	
Format	single tag removal	Weekly posts
SPARQL Update	18.7	1707
Tails Changeset	37.4	2977
GUO	31.9	1820
GRUF	18.0	1831
Sesame Transactions	21.9	2764

+ Conclusion



■ 技術の様々な局面の比較

- ❖ 議論: 分散組織環境での効果的なRDF dataの同期
 - ✓ ネットワークの観点でパブリッシャからupdateが最適
 - ✓ オーバヘッドはハブに移動
 - ✓ SPARQL Update1.0が最も効果的で柔軟
 - ◇ 良い粒度の記述
 - ◇ CRUDオペレーションも可能
 - ◇ 低いネットワークの利用

+ まとめと考察



- 今回のユースケース: Real-time integration for Synchronization
 - ❖ ネットワーク利用率
 - ❖ データ連携におけるPush/Pullの必要性
- 考察
 - ❖ 論文と本研究の視点
 - ✓ 論文: データ連携 = 情報の共有(eg. Wiki, blog, microblog)
 - ✓ 本研究: データ連携 = サービスのデータのやりとり
 - ⇒ 技術にとらわれずにもっと多くのデータ連携を考えられる?
- 学んだこと
 - ❖ データ連携におけるネットワーク利用率考慮 (粒度が低すぎる?)

+ ユースケース探索



■ 視点

- ❖ Linked Data integration
- ❖ Service integration used linked data

■ 読んだ論文

- ❖ Using Linked Data for Systems Management
- ❖ Using OWL in Data integration
- ❖ Toward the Next Wave of Service
- ❖ Towards Linked Open Service and Process

⇒ 前回のシーケンスを適用できるユースケースではなかった

⇒ 視点の変更を行う

+ 今後の課題



■ ユースケースの決定

❖ ユースケース探索

✓ 観点: クラウドサービス連携の論文

◇ Linked Dataを用いていないユースケースに適用すべき?

✓ 論文を深く読まない

❖ クラウドサービスとその他の技術(Blog, wiki)の連携の考察

+ 参考文献



- SIOC project, <http://sioc-project.org/>
- SPARQL Updateと読み書きウェブとTabulator, http://s-web.sfc.keio.ac.jp/committee2009/0626-read_write_web-kennyluck/