



*Empoderando vidas.  
Fortalecendo nações.*

## **Projeto BRA/12/018 - Desenvolvimento de Metodologias de Articulação e Gestão de Políticas Públicas para Promoção da Democracia Participativa**

### **Produto 05 - Proposta de regras de extração de conteúdos da API do portal e suas ferramentas para alimentação de eventual/hipotética base/nuvem de conhecimento de participação social**

potencializando leituras focadas em incidência e participação social nas políticas públicas, com propostas de códigos

Renato Fabbri



**Secretaria-Geral da Presidência da República**

# **Produto 05 - Proposta de regras de extração de conteúdos da API do portal e suas ferramentas para alimentação de eventual/hipotética base/nuvem de conhecimento de participação social**

---

**Contrato n. 2013/000566**

**Objeto da contratação: Aporte de conhecimentos e tecnologias para especificação de vocabulário e ferramentas assistidas que utilizam processamento de linguagem natural e análise de redes complexas para o conteúdo do portal da participação social.**

Valor do produto: R\$ 21,600 (vinte e um mil e seiscentos reais)

Data de entrega: 12 de Novembro de 2014

Nome d@ consultor(a): Renato Fabbri

Nome d@ supervisor(a): Ricardo Poppi



Fabbri, Renato

Proposta de regras de extração de conteúdos da API do portal e suas ferramentas para alimentação de eventual/hipotética base/nuvem de conhecimento de participação social: potencializando leituras focadas em incidência e participação social nas políticas públicas, com propostas de códigos / 2014.

Total de folhas: 17

Supervisor(a): Ricardo Poppi

Secretaria: SNAS

Secretaria-Geral da Presidência da República

Palavras-chave: reconhecimento de padrões, redes complexas, processamento de linguagem natural, web semântica, participação social.



Esta obra é licenciada sob uma licença Creative Commons - Atribuição-NãoComercial. 4.0 Internacional.



# Sumário

<b>1 Introdução</b>	<b>7</b>
1.1 Contexto e importância da consultoria . . . . .	7
1.2 Contexto e importância do Produto . . . . .	7
1.2.1 Objetivos . . . . .	7
1.2.2 Resultados esperados . . . . .	7
1.2.3 Caráter inovador . . . . .	7
<b>2 Desenvolvimento</b>	<b>7</b>
2.1 Etapas de desenvolvimento anteriores a este produto . . . . .	7
2.1.1 Sistematização ontológica da participação online . . . . .	7
2.1.2 Triplificação dos dados do participa.br . . . . .	7
2.1.3 Levantamento do endpoint SparQL . . . . .	8
2.1.4 Análises iniciais, modelos . . . . .	8
2.1.5 Sistema de recomendação de participante e recursos . . . . .	8
2.2 Etapas de desenvolvimento deste produto . . . . .	8
2.3 Justificativa, descrição detalhada e formas de aplicação do método . . . . .	8
2.4 Justificativa, descrição detalhada e acesso das fontes . . . . .	8
<b>3 Resultados alcançados</b>	<b>8</b>
3.1 Usos dos resultados . . . . .	8
<b>4 Conclusão</b>	<b>8</b>
4.1 Comentários, sugestões, recomendações . . . . .	8
4.2 Impacto do Produto para a elaboração, gestão e/ou avaliação de políticas públicas de participação social . . . . .	8
4.3 Impacto no público-alvo das políticas públicas a que se refere . . . . .	8
<b>5 Agradecimentos</b>	<b>8</b>
<b>A Ontologias de instâncias participativas online potencialmente relacionáveis ao participa.br</b>	<b>12</b>
A.1 Ontologia do AA (Ontologiaa) . . . . .	12
A.2 Ontologia da Cidade Democrática (OCD) . . . . .	14
<b>B Revisão da OPA</b>	<b>15</b>
<b>C Revisão da Triplificação do Participa.br</b>	<b>15</b>



<b>D Ontologia e Vocabulário da Biblioteca Social (OBS e VBS)</b>	<b>15</b>
D.1 PNPS . . . . .	15
D.2 Entrevistas . . . . .	15
D.3 Relatório de implementação das contribuições do Workshop dia 20/Out/2014, sobre a biblioteca (semântica de participação) social . . . . .	15
D.4 Materiais enviados pela equipe para referência . . . . .	15
<b>E Utilização dos dados linkados</b>	<b>15</b>
E.1 Pubby . . . . .	15
E.2 Endpoint SparQL . . . . .	17
E.3 Webprotege . . . . .	17



## Resumo

Este documento descreve o quinto produto.

**Palavras-chave:** reconhecimento de padrões, redes complexas, processamento de linguagem natural, web semântica, participação social.



# 1 Introdução

## 1.1 Contexto e importância da consultoria

descrever o objetivo GERAL da consultoria e como este Produto específico está contextualizado dentro do objetivo final da contratação

## 1.2 Contexto e importância do Produto

### 1.2.1 Objetivos

### 1.2.2 Resultados esperados

### 1.2.3 Caráter inovador

destacar como este trabalho poderá contribuir suprir uma lacuna de conhecimento e/ou para desenvolver determinada a capacidade institucional da SG/PR.

# 2 Desenvolvimento

espaço onde o consultor vai construir suas ideias. O consultor tem a liberdade para organizá-lo em tópicos, itens e sub-itens.

demonstrar que o produto entregue corresponde ao que foi solicitado no termo de referência, por meio de: 4. Análise sobre os resultados esperados na etapa de planejamento do Produto e os resultados alcançados ao final do Produto.

## 2.1 Etapas de desenvolvimento anteriores a este produto

Descrição detalhada das etapas de desenvolvimento do Produto

### 2.1.1 Sistematização ontológica da participação online

Através de estudos e reuniões presenciais e online, a Ontologia de Participação Social (OPS) foi revisada [1] e a Ontologia do Participa.br (OPA) foi feita [2].

### 2.1.2 Triplificação dos dados do participa.br

Feito um script para triplificar os dados do Participa.br, ou seja, para o enriquecimento semântico e escrita em RDF dos dados em Postgresql da instância Noosfero do Participa.br [3].



### **2.1.3 Levantamento do endpoint SparQL**

Para uso dos dados triplificados, pode-se recorrer a diversos métodos de leitura e disponibilização. Um método-chave é a disponibilização dos dados rdf (*triple store*) em um *endpoint sparql*. Para os fins de testes, pesquisa e usos leves, está disponibilizado um endpoint SparQL em servidores da USP [4].

### **2.1.4 Análises iniciais, modelos**

Análises dos dados do participa.br foram abertas no IPython Notebook, com ênfase no texto produzido e nas redes formadas [5].

### **2.1.5 Sistema de recomendação de participante e recursos**

## **2.2 Etapas de desenvolvimento deste produto**

## **2.3 Justificativa, descrição detalhada e formas de aplicação do método**

## **2.4 Justificativa, descrição detalhada e acesso das fontes**

## **3 Resultados alcançados**

### **3.1 Usos dos resultados**

## **4 Conclusão**

retomar as ideias trabalhadas ao longo do Produto e fazer uma análise sobre as mesmas.

### **4.1 Comentários, sugestões, recomendações**

### **4.2 Impacto do Produto para a elaboração, gestão e/ou avaliação de políticas públicas de participação social**

### **4.3 Impacto no público-alvo das políticas públicas a que se refere**

## **5 Agradecimentos**

O consultor Renato Fabbri agradece ao Joenio Costa pelo template em L<sup>A</sup>T<sub>E</sub>X para os produtos. Agradece à Daniela Feitosa pela reunião para demanda de recomendação de perfis. Agradece



*Empoderando vidas.  
Fortalecendo nações.*

aos supervisores do trabalho realizado em torno do participa.br: Ricardo Poppi e Ronald Costa. Agradece ao labMacambira.sf.net e todas as comunidades de software e cultura livre que compõe esta contribuição.



## Referências

- [1] “Ontologia de participação social,” <http://tinyurl.com/p2doueu>.
- [2] “Ontologia do participa,” <http://tinyurl.com/lcccwop>.
- [3] “Especificação da triplificação dos dados do participa.br,” <http://tinyurl.com/k74z3yl>.
- [4] *Endpoint SparQL com dados do Participa.br.* <http://200.144.255.210:8082/>.
- [5] *Repositório Git do produto 3: documento e scripts.* <https://github.com/ttm/pnud3>.



## Abreviações e jargão

**RC:** Redes Complexas

**PLN:** Processamento de Linguagem Natural

**OPS:** Ontologia de participação Social

**OPA:** Ontologia do Participa.br

**MMISSA:** Monitoramento Massivo e Interativo da Sociedade pela Sociedade para Aproveitamento

**AARS:** A Análise de Redes Sociais

**MyNSA:** Monitoring yields Natural Streaming and Analysis

**PNPS:** Plano Nacional de Participação Social

**RDF:** Resource Description Framework

**HTTP:** Hypertext Transfer Protocol

**SPARQL:** Simple Protocol and RDF Query Language

**endpoint SPARQL:** ponto de acesso, geralmente HTTP, a dados em RDF via buscas em SPARQL.

**Participa.br:** Portal federal de participação social.

**IPython Notebook:** instância online para rodar scritps Python

**Meteor:** arcabouço para páginas reativas e com funcionamento distribuído.

**D3js:** biblioteca de visualização de dados.



## A Ontologias de instâncias participativas online potencialmente relacionáveis ao participa.br

### A.1 Ontologia do AA (Ontologiaaa)

Como uma forma de integrar o Participa.br em uma nuvem de conhecimento participativo, foi levantada a Ontologiaaa, exposta na Figura 1. O AA é uma técnica de compartilhamento de processos usada principalmente no labMacambira.sf.net. A simplicidade das implementações atuais, e a pertinência do registro e compartilhamento de processos, fizeram com que esta fosse o primeiro desenvolvimento efetivo deste último produto.

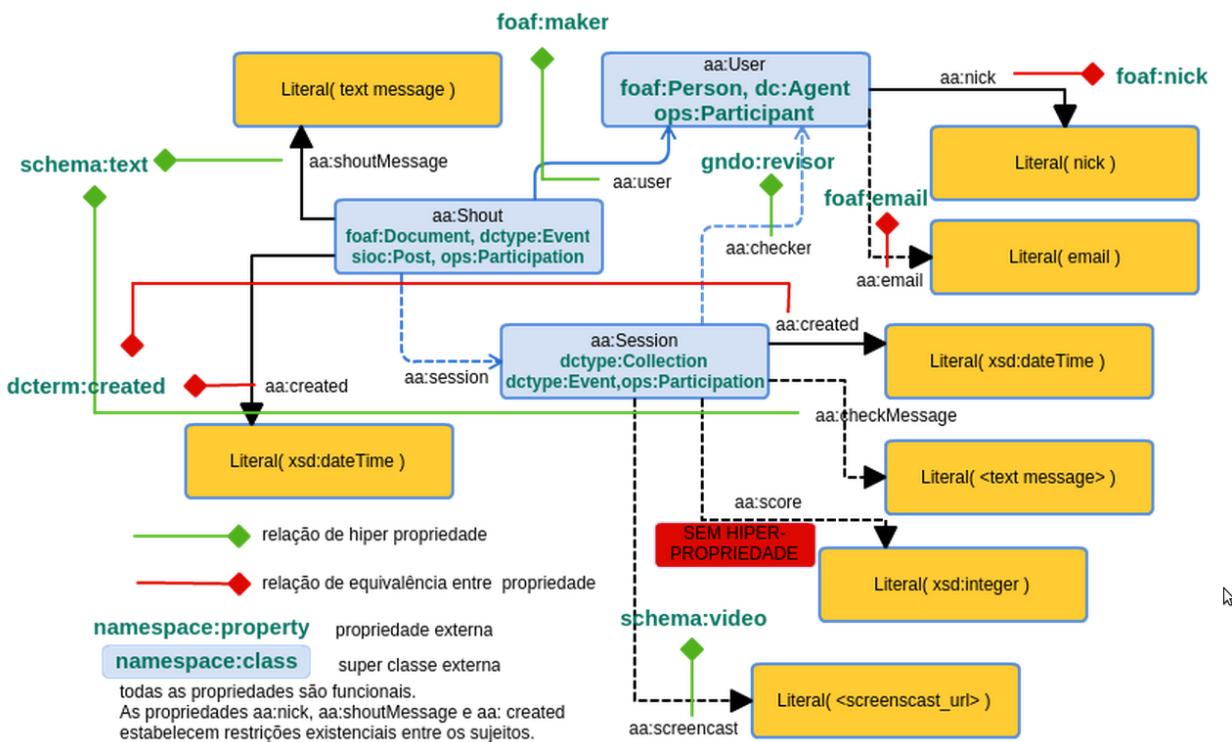


Figura 1: Ontologia do AA, com suas classes, propriedades, literais, e classes e propriedades externas usadas para relacionar os dados do AA aos do participa e de toda nuvem LOD.

O tamanho reduzido da ontologia permitiu que vários testes fossem feitos. Em especial, com a ontologiaaa foi reestabelecida a arquitetura de ontologia com uso de um namespace interno (no caso <http://purl.org/socialparticipation/aa/>) e inferências para contemplar outros namespaces.



As inferências foram testadas com o jena/fuseki, com bons resultados. Tanto as inferências relacionadas às hiperônimas (superclasses e super propriedades, diretamente do rdfs) quanto inferências mais elaboradas (ligadas ao padrão OWL) foram satisfatórias. O revés é que qualquer query SparQL que demora milissegundos, mesmo que não envolva inferências para sua resposta, demora segundos quando há uma máquina de inferências ativa. A solução, portanto, parece ser ainda de realizar estas inferências offline e disponibilizar todas as triplas resultantes no endpoint.

Todos os desenvolvimentos desta ontologia e a triplificação de dados do AA em MySQL e MongoDB estão em: <https://github.com/ttm/aa01/tree/master/rdf>. Estes dados estão disponíveis no endpoint sparql (fuseki/jena) para uso conforme `scriptipython`. Há interfaces úteis para explorar/expor os dados ligados ao AA. Em especial, estão derreferenciáveis, como na Figura 2.

The screenshot shows a Google Chrome window titled 'Ontologias de participação social - Google Chrome'. The address bar contains the URL <http://purl.org/socialparticipation/aa/Shout%2339722>. The main content area displays the RDF triples for this entity:

Property	Value
aa:created	▪ 2012-02-21T18:16:15 ()
aa:session	▪ aa:Session%231479
aa:shoutMessage	▪ alert indo para mapeamentoSEC2
rdf:type	▪ aa:Shout
aa:user	▪ aa:User%23bitanoma
aa:valid	▪ false ()

Below the table, there are links to 'As Turtle', 'As RDF/XML', 'Browse in Disco', 'Browse in Tabulator', and 'Browse in OpenLink Browser'.

Figura 2: Mensagem (shout) do AA derreferenciado. Cada mensagem do AA recebe uma URI, assim como cada sessão e cada usuário. Estes três conceitos são instânciados com URIs dedicadas, e relacionadas via ainda outras URIs. Por fim URIs especificam relações entre instâncias destes conceitos e os dados.



## A.2 Ontologia do Cidade Democrática (OCD)

Outra instância participativa considerada prioritária pelo consultor para integração aos dados participativos linkados, e contemplada neste trabalho, foi o portal Cidade Democrática. Este portal possui grande complexidade e abundância de dados e conceitos. Assim, esta empreitada contrastou com a da Ontologia descrito no Apêndice A.1.

Com a grande complexidade das tabelas e dados, foi feita uma decupagem do banco de dados (disponibilizada em <https://github.com/ttm/ocd/blob/master/decupagemBD.txt>) e uma triplificação destes dados (script em: <https://github.com/ttm/ocd/blob/master/triplificaCD.py> e triplas resultantes em <https://github.com/ttm/ocd/blob/master/cdTriplestore.rdf.tar.gz>).

Embora os trabalhos de decupagem do banco e de triplificação dos dados sejam expressivos, o ponto alto desta empreitada foi a gênese de um método de levantamento de ontologia orientado aos dados. Este método é extremamente útil para qualquer portal que queira representar seus dados como triplas RDF e uma ontologia. O processo é o seguinte:

1. Todos os dados de interesse são triplificados com namespace interno, conforme: <https://github.com/ttm/ocd/blob/master/triplificaCD.py>.
2. Os dados triplificados são disponibilizados em um endpoint sparql para levantamento da ontologia com base nas triplas produzidas (endpoint em: <http://200.144.255.210:8082/cd/query>).
3. Um script é construído, no qual os dados triplificados são usados para observação das estruturas ocorrentes, conforme <https://github.com/ttm/ocd/blob/master/OCD.py>. Principalmente:
  - São observadas todas as classes ocorrentes.
  - São observadas todas as propriedades ocorrentes.
  - São feitas imagens de cada classe e de cada propriedade, com os elementos imediatamente relacionados a eles, como nas Figuras ?? e ??.



- B Revisão da OPA**
- C Revisão da Triplificação do Participa.br**
- D Ontologia e Vocabulário da Biblioteca Social (OBS e VBS)**
  - D.1 PNPS**
  - D.2 Entrevistas**
  - D.3 Relatório de implementação das contribuições do Workshop dia 20/Out/2014, sobre a biblioteca (semântica de participação) social**
  - D.4 Materiais enviados pela equipe para referência**
- E Utilização dos dados linkados**

As ontologias mais importantes para este trabalho precisam ser observáveis e anotáveis com facilidade por não especialistas.

Uma primeira opção é aproveitar os dados/tripas disponíveis no endpoint sparql já aberto com uma instância Fuseki/Jena. Este endpoint pode ser acessado via linguagens de scripting para prototipação rápida, como JavaScript ou Python, fornecendo interfaces gráficas e web para navegação e análise. Estas possibilidades estão desenvolvidas nos produtos 2 e 3 desta mesma consultoria [?, ?].

### **E.1 Pubby**

Com desenvolvimento recente no github <https://github.com/cygri/pubby>, é talvez o navegador de dados mais conhecido. Parece ser projeto do dig (grupo do Berners-Lee no MIT). Os testes mostraram que as consultas sparql demoravam demais para o montante de dados triplificados, portanto foram importados os rdfs. Para facilitar, o arquivo de configuração do pubby está em <https://github.com/ttm/vocabulario-participacao/blob/master/auxiliar/config.ttl>.

A url permanente <http://purl.org/socialparticipation/> é redirecionada para a instância do pubby, em: <http://200.144.255.210:8081/tpubby/page/>. Desta forma, *todos* os conceitos da VBS e classes da OBS podem ser derreferenciados, como na Figura 4.



*Empoderando vidas.  
Fortalecendo nações.*



Figura 3: Diagrama de Ouvidorias desenhado com o acompanhamento de especialista (Lígia).



The screenshot shows a Google Chrome browser window with the title "Instância de participação social | Ontologias de participação social - Google Chrome". The address bar contains the URI <http://purl.org/socialparticipation/obs/ParticipationInstance>. The page content is an RDFa representation of the concept:

Property	Value
rdfs:label	▪ Instância de participação social (pt)
is rdfs:subClassOf of	▪ obs:Commission ▪ obs:Conference ▪ obs:Council ▪ obs:OmbudsmanAgency ▪ obs:Ombudsmen
rdfs:subClassOf	▪ obs:ParticipationInstanceOrMechanism
rdf:type	▪ owl:Class

Below the table, there are links: "As Turtle", "As RDF/XML", "Browse in Disco", "Browse in Tabulator", and "Browse in OpenLink Browser".

Figura 4: Derreferenciamento do conceito de Instância de participação através da URI <http://purl.org/socialparticipation/obs/ParticipationInstance>.

## E.2 Endpoint SparQL

Os dados da OBS, VBS, Participab, AA e OCD estão disponíveis via endpoint SparQL, conforme o uso explicitado nos scripts do INotebook <http://200.144.255.210:8003/>.

## E.3 Webproteg

Permite que as ontologias e vocabulários estejam online e comentáveis.