

# Sistema de Recomendação do Portal gov.br

Rafael Odon de Alencar<sup>1</sup>, Marcelo Pita<sup>1</sup>, Ronaldo Agra<sup>1</sup>

<sup>1</sup>SERPRO – Serviço Federal de Processamento de Dados  
Brasília – DF – Brasil

{rafael.odon, marcelo.pita, jose-ronaldo.souza}@serpro.gov.br

**Abstract.** *This paper addresses the development of the recommender system of the Brazilian Government Portal (gov.br), which enhances citizen's experience by offering interesting digital services. Architectural decisions and implementation details regarding the proposed solution, which is based on information retrieval techniques, are discussed. Experimental results indicate that the proposed solution is most cost-effective when compared to a traditional matrix factorization-based recommendation technique.*

**Resumo.** *Este artigo aborda a construção do sistema de recomendação do Portal gov.br, que melhora a experiência do cidadão ao oferecer serviços digitais de seu interesse. Decisões arquiteturais e detalhes de implementação da solução proposta, baseada em técnicas de recuperação de informação, são discutidas. Resultados experimentais indicam que a solução apresenta a melhor relação custo-benefício em comparação com uma técnica tradicional de recomendação baseada em fatoração de matrizes.*

## 1. Introdução

Com o objetivo de melhorar a relação entre cidadão e Estado, tornando-a mais simples e direcionada às necessidades dos usuários de serviços públicos, em 2019 foi instituído o projeto *gov.br*. Trata-se de uma estratégia do Governo do Brasil para reunir, em um único local, todos os serviços e informações sobre a atuação dos diversos órgãos do governo federal, possibilitando aos cidadãos agilidade na sua interação com o Estado. Neste contexto, foi criado o Portal gov.br, como forma de unificar e padronizar a oferta de serviços públicos à sociedade brasileira [GOVERNO DO BRASIL - A, 2022].

A partir da implantação do Portal gov.br, esta oferta de serviços cresceu exponencialmente, saindo de 1.622 serviços digitais, no início do ano de 2019, e atingindo 4.905 serviços digitais de órgãos da administração pública federal nos dias atuais [GOVERNO DO BRASIL - B, 2022]. A centralização dos serviços permitiu aos cidadãos encontrar e acessar aqueles desejados com mais facilidade. Entretanto, ainda não é possível ter uma visão clara e simplificada da interdependência entre os serviços, muitas vezes complexa. Com isso, apesar do crescimento na quantidade de acessos ao portal, a conversão em solicitações de serviço não cresceu no mesmo ritmo.

Assim sendo, como melhorar a experiência do usuário do Portal gov.br, quando o mesmo está buscando por serviços digitais, de forma que sejam ofertadas também opções de serviços correlacionados ao buscado inicialmente? Indo mais adiante, mesmo que o usuário não esteja buscando um serviço explicitamente, seria possível o Portal gov.br ser proativo e oferecer serviços com base no seu histórico de navegação? E, por fim, como trazer mais clareza ao entendimento sobre a interdependência entre serviços digitais públicos, de tal modo que a oferta do portal tenha relevância? Questões como essas podem ser tratadas com o apoio de sistemas de recomendação [Resnick e Varian

1997]. As informações registradas sobre históricos de navegação dos usuários no portal, além de serem uma oportunidade para provisão de melhores recomendações de serviços, ajudam a clarificar as suas dependências.

O objetivo deste trabalho é apresentar a situação problema que culminou na prospecção, implementação e implantação de um sistema de recomendação para serviços do Portal gov.br, bem como discutir a solução proposta e resultados alcançados. O artigo está estruturado da seguinte forma: na Seção 2 é apresentada uma revisão bibliográfica; a Seção 3 detalha a solução proposta de sistema de recomendação para o Portal gov.br; nas Seções 4 e 5 são apresentados e discutidos os resultados experimentais do trabalho; por fim, na Seção 6, as conclusões e trabalhos futuros.

## **2. Sistemas de recomendação**

Sistemas de recomendação são sistemas de informação criados com a finalidade de sugerir itens – de qualquer natureza, *e.g.* livros, filmes, notícias, serviços, dentre outros – para seus usuários [Resnick e Varian 1997][Ricci *et al.* 2011][Aggarwal 2016]. Um bom sistema de recomendação deve sugerir, portanto, itens que são de interesse dos usuários, contemplando a especificidade e a variedade de suas preferências [Ricci *et al.* 2011]. Atualmente, estão presentes em uma grande variedade de contextos do dia a dia de usuários de serviços digitais. Por exemplo, em plataformas de *video streaming* [Gomez-Uribe e Hunt 2015][Ferreira *et al.* 2020], propagandas personalizadas exibidas em redes sociais [Gupta *et al.* 2020], sites de viagens [Logesh *et al.* 2018], *audio streaming* [Semerci *et al.* 2018], e-commerce [Linden *et al.* 2003], dentre outros.

### **2.1. Principais tipos de motores de recomendação**

Motores de recomendação podem ser implementados usando diferentes estratégias e técnicas [Aggarwal 2016]. Destacam-se duas estratégias como as mais usadas para a construção de motores de recomendação: (i) *baseada em conteúdo*; e (ii) *filtragem colaborativa*. A recomendação baseada em conteúdo utiliza as preferências do usuário por itens para prover sugestões de outros itens que são similares aos já manifestados como de seu interesse. Por outro lado, a recomendação por filtragem colaborativa utiliza as preferências dos usuários para definir graus de similaridade entre eles, e nesse caso o motor de recomendação irá sugerir itens que ainda não lhe foram apresentados, mas que se revelaram como de interesse nas preferências de outros usuários similares. Na prática, motores de recomendação costumam usar *estratégias híbridas* [Lu *et al.* 2013].

### **2.2. Sistemas de recomendação para Governo**

As ondas de transformação digital que tem informatizado serviços públicos em Governos de todo o mundo cria um cenário de sobrecarga informacional para cidadãos e empresas, que, diante de processos de caráter naturalmente burocráticos, enfrentam dificuldades práticas no acesso e realização desses serviços [Sun *et al.*, 2021]. Sistemas de recomendação podem, nesse cenário, se tornar ferramentas efetivas que ajudam a disponibilizar aos usuários os serviços públicos de que realmente necessitam.

[Al-Hassan *et al.* 2009] propõem um arcabouço conceitual, chamado Pe-Gov, para recomendação personalizada de serviços governamentais centrada no cidadão. O arcabouço é baseado no uso de ontologias como forma de estruturar e integrar conhecimentos específicos de domínio, usando informações e *feedbacks* de usuários autenticados. [Lu *et al.*, 2010] apresenta o sistema de recomendação *BizSeeker*, uma implementação *government-to-business* baseada em uma abordagem híbrida que ajuda pequenas e médias empresas australianas a encontrarem novos parceiros de negócio. Uma evolução desse último trabalho, que incorpora características *fuzzy* ao motor de recomendação, é o *Smart BizSeeker* [Lu *et al.*, 2013]. [Ayachi *et al.*, 2016] também propõem um arcabouço conceitual para recomendação de serviços governamentais, usando como exemplo o portal de serviços do Governo de Quebec<sup>1</sup>, em que serviços personalizados são ofertados aos cidadãos autenticados com base em diferentes fontes de dados. Mais recentemente, [Sun *et al.*, 2021] apresentaram uma nova técnica de recomendação de serviços de governo (distrito de Wuhou - Chengdu, China) que leva em consideração as interações negativas dos usuários em filtragem colaborativa.

No Brasil, [Macedo *et al.*, 2020] propõem uma ferramenta de recomendação de visualizações adequadas para conjuntos de dados abertos com o intuito de promover mais transparência pública. Um modelo conceitual geral para recomendações de serviços públicos ao cidadão é proposto em [Moisinho *et al.*, 2021], que contempla aspectos éticos no trato de dados pessoais ao incorporar a noção de auto perfil.

### **3. Recomendação de serviços no Portal gov.br**

O Portal gov.br centraliza a oferta de serviços públicos de diferentes órgãos da administração pública federal. Cada serviço possui sua respectiva página *Web*, onde o visitante obtém detalhes sobre seu funcionamento, podendo iniciar o acionamento do serviço caso ele seja digital. Alguns exemplos de serviços são: atualizar o CPF; consultar restituição de imposto de renda; solicitar registro de pescador amador e obter o certificado nacional de vacinação COVID-19.

#### **3.1. Principais desafios**

A criação de um sistema de recomendação de serviços no Portal gov.br foi possível a partir do histórico de acesso aos endereços internos do portal pelos usuários. Os acessos são anônimos, ou seja, não possuem qualquer identificação do cidadão, porém são associados a um identificador (ID) de cliente *Web*, que agrupa todos os acessos feitos por um mesmo navegador em um mesmo dispositivo. Este identificador possibilita manter algum nível de rastreamento entre os acessos feitos por um mesmo visitante, viabilizando análises do tipo: quem acessou A e B também acessou C. Apesar disso, ainda existem alguns desafios:

---

<sup>1</sup> Portal do Governo de Quebec: <https://www.quebec.ca/>

- Um mesmo cidadão pode acessar serviços em navegadores e dispositivos diferentes, em locais diferentes, ou mesmo limpar o *cache* do navegador que faz maior uso, prejudicando o rastreamento de serviços acessados por ele;
- Em alguns casos, um cidadão pode acessar o serviço digital diretamente, sem ter navegado pelo Portal gov.br para acioná-lo, perdendo assim a rastreabilidade;
- O ID de Cliente Web pode agrupar acessos de mais de um cidadão quando se trata de um navegador compartilhado entre diferentes pessoas.

Observou-se que cerca de 75% dos ID de Cliente *Web* estavam associados a uma única visita. Porém, o grande volume de informações de acessos – cerca de 10 milhões de visualizações de páginas por dia – ainda continuaria a oferecer um grande valor para o cálculo das recomendações, tornando desprezíveis as inconsistências apontadas.

### 3.2. Solução desenvolvida

O diagrama apresentado na Figura 1 destaca os principais elementos arquiteturais do sistema de recomendação proposto para o Portal gov.br, sendo eles:

1. **Módulo de rastreamento dos acessos** (front-end): mantém um histórico local dos acessos recentes de um visitante e alimenta o histórico de acessos com dados que permitem rastrear acessos oriundos de um mesmo cliente *Web*;
2. **Pipeline de tratamento dos dados** (back-end): ingere periodicamente o histórico de acessos, realizando transformações que apoiam a recomendação;
3. **Motor de recomendação** (back-end): calcula a recomendação com base no aprendizado obtido no histórico de acessos;
4. **API de recomendação** (back-end): *endpoint* HTTP escalável que suporta milhares de visitantes simultâneos com ajuda de estratégias de *cache*;
5. **Módulo de recomendação** (front-end): consome a API de recomendação para montar a caixa de recomendações exibida no portal com base nos acessos recentes realizados pelo usuário que ficam salvos no histórico local.

O ponto de partida da solução é a ingestão periódica do histórico de acessos a serviços. Esse histórico de acessos é transformado em uma estrutura contendo o *ID de Cliente Web* como chave e os serviços distintos já acessados por ele como valor (Tabela 1.A). Novos acessos são contabilizados de forma incremental. O motor de recomendação desenvolvido é baseado em técnicas de recuperação de informação [Baeza-Yates e Ribeiro-Neto 1999], ou seja, utiliza buscas indexadas em um índice invertido. Para tanto, modela-se o “carrinho de visitas” de cada visitante como um documento indexado, onde cada serviço já visitado é uma *keyword* diferente. Isso permite criar um índice invertido (Tabela 1.B), utilizado pelo motor de busca para encontrar todos os documentos que possuem um serviço específico (Tabela 1.C). No subconjunto de documentos que atendem à busca por uma determinada *keyword*, ou serviço, é possível analisar a quantidade de coocorrência de outras *keywords*, ou serviços (Tabela 1.D).

No exemplo demonstrado na Tabela 1.D, se fosse desejado recomendar 2 serviços relacionados ao A, poderíamos recomendar C e B, sendo C mais relevante que B por ter ocorrido mais vezes em conjunto com A. Uma vantagem dessa estratégia de

recomendação baseada em recuperação de informação é que a recomendação é calculada *online*, em tempo de consulta, se apoiando nas estruturas eficientes dos índices de consulta. Isso permite que novos documentos sejam indexados periodicamente (novos trechos do histórico de acesso) refletindo na recomendação tão logo o índice invertido seja atualizado. Na prática, equivale a uma filtragem colaborativa baseada em itens cujo ajuste do modelo é feito de forma incremental.

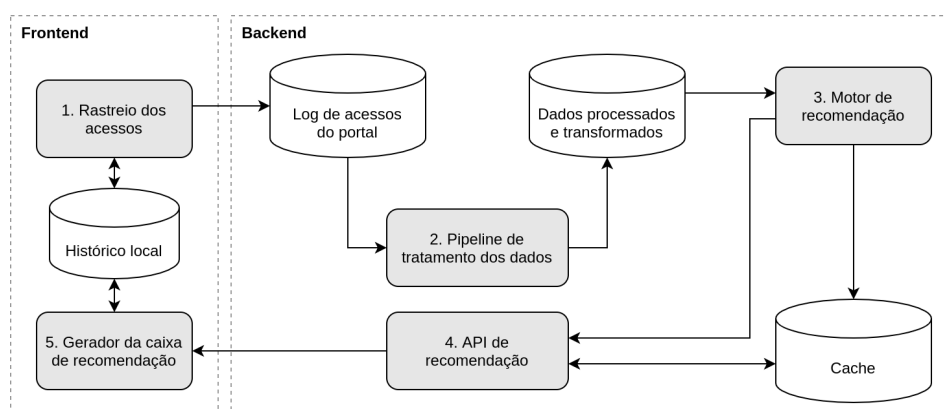


Figura 1. Arquitetura em alto nível do sistema de recomendação do Portal gov.br.

Tabela 1. Dados e transformações realizadas no histórico de acessos de usuário: (a) serviços visitados por ID Cliente Web; (b) índice invertido; (c) agregação para um certo serviço; (d) frequência de cocorrência de serviços na agregação.

| (a)            |                    | (b)               |                       | (c)  |         | (d)     |        |
|----------------|--------------------|-------------------|-----------------------|------|---------|---------|--------|
| ID Cliente Web | Serviços Visitados | Keyword (serviço) | Docs (ID Cliente Web) | Docs | Keyword | Keyword | Ocorr. |
| #1             | A, B, C            | A                 | #1, #2, #3            | #1   | A, B, C | A       | 3      |
| #2             | A                  | B                 | #1                    | #2   | A       | C       | 2      |
| #3             | A, C               | C                 | #1, #3                | #3   | A, B    | B       | 1      |
| #4             | D                  | D                 | #4                    |      |         |         |        |

O motor de recomendação construído é exposto através de uma API HTTP. A medida que o usuário navega pelas páginas de serviços do portal, a recomendação é atualizada dinamicamente. Para gerar essa recomendação contextualizada com a navegação recente, o portal mantém no histórico local do navegador os últimos serviços distintos acessados pelo visitante. Esses serviços são repassados à API do motor de recomendação que, por sua vez, utiliza-os como critério de busca. Uma questão observada nas recomendações geradas por essa estratégia foi que, em determinados momentos, serviços do histórico recente mais populares podem levar a uma predominância de um assunto em detrimento de outros menos populares, que ficariam de fora da recomendação. Para resolver essa questão, adotou-se então uma estratégia em que são feitas buscas separadas, uma para cada serviço distinto do histórico, mesclando-se os resultados ao final de forma balanceada (Algoritmo 1).

Consultas separadas permitem manter em *cache* resultados separados para cada um dos 5 mil serviços do Portal gov.br. Já para manter em *cache* o resultado das

consultas com todos os serviços ao mesmo tempo, demanda-se um espaço bem maior (a combinação simples dos quase 5 mil serviços tomados de 3 em 3, no caso de um histórico de tamanho 3). Diversos níveis de *cache* foram implementados para garantir um bom desempenho da solução, tendo em vista o alto volume de acessos diários.

|    |  |
|----|--|
| 1  | <b>Função</b> MesclarRecomendações(recomendaçõesServiços[[]], tamanhoSaida):                         |
| 2  | saída = Conjunto(vazio)  |
| 3  | <b>Enquanto</b> tamanho(saída) < tamanhoSaida <b>faça</b> :  |
| 4  | <b>Para cada</b> recomendações <b>em</b> recomendaçõesServiços <b>faça</b> :                         |
| 5  | <b>Se</b> tamanho(recomendações) > 0 <b>então</b> :  |
| 6  | mesclado.adicionar(removeDoTopo(recomendações))  |
| 7  | total = somar(tamanho(recomendações) <b>para cada</b> recomendações <b>em</b> recomendaçõesServiços) |
| 8  | <b>Se</b> total == 0 <b>então</b> :  |
| 9  | <b>Parar</b>   |
| 10 | <b>Retornar</b> saída  |

**Algoritmo 1: Mesclar recomendações distintas para serviços em um único resultado.**

## 4. Resultados Experimentais

Neste trabalho foram realizados 3 tipos de experimentos visando avaliar e comparar os resultados obtidos em estratégias distintas de recomendação. Foram eles:

- Experimento 1: comparação preliminar entre uma solução tradicional de recomendação baseada em fatoração de matrizes e a solução proposta de recomendação baseada em técnicas de recuperação de informação;
- Experimento 2: comparação de diferentes métodos de cálculo de relevância na solução proposta;
- Experimento 3: comparação de diferentes formas de composição de resultados na solução proposta.

A fim de experimentar diferentes estratégias de recomendação, foi definida uma abordagem onde o modelo de recomendação é ajustado com dados históricos de acesso para uma dada janela de tempo (conjunto de treinamento). Em seguida são observados, na janela imediatamente subsequente, quais serviços presentes na recomendação fornecida foram de fato acessados pelos usuários (conjunto de teste, com dados não vistos no treinamento). Os conjuntos de dados utilizados estão descritos na Seção 4.1. A Seção 4.2 descreve as métricas de avaliação utilizadas, enquanto as Seções 4.3 a 4.5 descrevem os experimentos e respectivos resultados obtidos.

### 4.1. Conjuntos de dados

Os motores de recomendação comparados nos experimentos 1 e 2 foram treinados com dados reais de acesso às páginas de serviços do Portal gov.br, obtidos entre os dias 07/01/2021 e 28/02/2021, e avaliados em uma base de acessos do portal referentes à primeira semana de março de 2021 (01/03/2021 a 07/03/2021). Ao todo foram selecionados, no histórico de acessos, cerca de *60 mil visitantes diferentes* que fizeram acessos a serviços tanto na janela de teste quanto na janela de treinamento. O experimento 3 usou dados históricos referentes a um período maior. Para este, o histórico de acesso de 01/11/2020 até 03/04/2020 (5 meses) foi utilizado como período

de treino. Já o histórico de acesso na janela de 04/04/2021 até 11/04/2021 foi utilizado como período de teste. Com isso, constaram 595.439 *visitantes distintos* que fizeram acesso tanto no período de teste quanto no período de treino da recomendação.

## 4.2. Métricas de Avaliação

Os modelos foram avaliados objetivamente através das métricas de precisão e revocação. Ambas são baseadas na quantidade de recomendações relevantes ( $R$ ) indicadas pelo sistema para cada usuário e entendidas, no contexto do Portal gov.br, como serviços que foram recomendados e que de fato foram visitados pelo usuário. As métricas são descritas a seguir.

**Precisão** – Fração de  $R$  em relação ao número total de recomendações  $N$  (Equação 1).

$$Prec@N = \frac{|R|}{|M|} \quad (1)$$

**Revocação** – Fração de  $R$  em relação ao número total de visitas do usuário  $|V|$ , considerando as top- $N$  recomendações (Equação 2).

$$Revoc@N = \frac{|R|}{|V|} \quad (2)$$

No experimentos foram calculadas precisão e revocação para cada visitante, a partir das quais foram calculadas as médias, para  $N \in [1, 15]$ .

## 4.3. Comparação de recomendação por fatoração de matrizes com estratégia de recomendação por recuperação de informação

Este experimento avalia e compara a capacidade preditiva de um motor de recomendação tradicional baseado em fatoração de matrizes (ALS) com a estratégia proposta de recomendação por meio de recuperação de informação (RecomRI).

**Configuração experimental** – Foi utilizada uma implementação em Python do algoritmo ALS (*alternating least squares*) [Koren e Bell 2015], disponível na API Spark (versão 2.4.4)<sup>2</sup>, com execução paralela em um cluster Hadoop com 32 núcleos de processamento (8 executores, cada um com 4 cores) e 256GB de RAM reservados. O algoritmo utilizou os seguintes parâmetros: número de fatores latentes (componentes): 100; parâmetro regulador de *feedback* implícito ( $\alpha$ ): 1.0; parâmetro de regularização ( $\lambda$ ): 0.0; número máximo de iterações do algoritmo: 30; número de recomendações geradas por usuário: 15.

A estratégia RecomRI foi implementada usando a ferramenta Elasticsearch versão 7.9.1<sup>3</sup>, single node com 2GB de RAM. Foram utilizados os seguintes parâmetros:

<sup>2</sup> Disponível: <https://spark.apache.org/docs/2.4.4/index.html> (Acesso: 29/03/2022)

<sup>3</sup> Disponível: <https://www.elastic.co/pt/downloads/past-releases/elasticsearch-7-9-1> (Acesso: 29/03/2022)

escore para cálculo dos termos significativos: JLH<sup>4</sup> (default); quantidade mínima de resultados para compor resposta: 2; número de acessos recuperados do histórico do usuário para criação da consulta: até 15 (mais recentes); número de recomendações geradas por usuário: 15.

**Resultados** – A Tabela 2 exibe os resultados de precisão e revocação médios para o ALS-100 (i.e. ALS com 100 fatores latentes) e RecomRI. Podemos perceber que, em termos absolutos, o RecomRI apresentou melhor desempenho médio que o ALS-100. As diferenças são mais expressivas em termos de revocação.

**Tabela 2. Resultados de precisão e revocação médias do ALS-100 e RecomRI para diferentes valores de N. Os melhores resultados estão em negrito.**

| N  | Precisão média ALS-100 | Precisão média RecomRI | Revocação média ALS-100 | Revocação média RecomRI |
|----|------------------------|------------------------|-------------------------|-------------------------|
| 1  | 0.425188               | <b>0.432672</b>        | 0.355030                | <b>0.363797</b>         |
| 2  | 0.269434               | <b>0.284269</b>        | 0.427258                | <b>0.453706</b>         |
| 3  | 0.198454               | <b>0.214005</b>        | 0.460042                | <b>0.498488</b>         |
| 4  | 0.156522               | <b>0.171434</b>        | 0.477229                | <b>0.523896</b>         |
| 5  | 0.129163               | <b>0.143387</b>        | 0.488492                | <b>0.542446</b>         |
| 6  | 0.110395               | <b>0.122900</b>        | 0.498645                | <b>0.554503</b>         |
| 7  | 0.096966               | <b>0.107921</b>        | 0.508758                | <b>0.565275</b>         |
| 8  | 0.086006               | <b>0.096312</b>        | 0.514411                | <b>0.574356</b>         |
| 9  | 0.077473               | <b>0.087047</b>        | 0.520172                | <b>0.581922</b>         |
| 10 | 0.070377               | <b>0.079401</b>        | 0.524072                | <b>0.588040</b>         |
| 11 | 0.064653               | <b>0.073016</b>        | 0.528670                | <b>0.593462</b>         |
| 12 | 0.059826               | <b>0.067737</b>        | 0.532665                | <b>0.599087</b>         |
| 13 | 0.055679               | <b>0.063194</b>        | 0.536420                | <b>0.604291</b>         |
| 14 | 0.052063               | <b>0.059225</b>        | 0.539335                | <b>0.608786</b>         |
| 15 | 0.048972               | <b>0.055774</b>        | 0.542908                | <b>0.613225</b>         |

#### 4.4. Comparação de diferentes métodos de cálculo de relevância no RecomRI

O Elasticsearch permite calcular os itens relacionados a uma consulta através de uma agregação, que pode ser feita apenas com contabilização da quantidade absoluta de coocorrências de itens no subconjunto consultado (vide Tabela 1). Porém, a ferramenta também dispõe de outros métodos de escore, que calculam a relevância de itens considerando outros fatores. Os seguintes métodos estão disponíveis: *JLH*, *JLH com Mutual Information*, *Chi Square*, *Google Normalized Distance (GND)*, *Percentage*. O objetivo deste experimento é comparar a qualidade da recomendação do RecomRI gerada através dos diversos métodos citados, ajudando a calibrar a solução.

**Configuração experimental** – Foram utilizadas as mesmas configurações da Seção 4.3, exceto a quantidade de recomendações geradas por usuário, que foi configurada para 6.

<sup>4</sup> Disponível: [https://www.elastic.co/guide/en/elasticsearch/reference/current/search-aggregations-bucket-significantterms-aggregation.html#\\_jlh\\_score](https://www.elastic.co/guide/en/elasticsearch/reference/current/search-aggregations-bucket-significantterms-aggregation.html#_jlh_score) (Acesso: 29/03/2022)



**Resultados** – Conforme resultados apresentados na Tabela 3, é possível observar que as diferentes estratégias comparadas possuem resultados muito próximos, com exceção da *Percentage*, que se mostra bastante inferior às demais. A estratégia que possui os melhores resultados de revocação é a de contagem absoluta das ocorrências.

**Tabela 3: Resultados comparativos de precisão e revocação para diferentes métodos de escore de relevância no RecomRI. Os melhores resultados estão em negrito.**

| Método                           | Precisão Média | Revocação Média |
|----------------------------------|----------------|-----------------|
| JLH                              | 0.1901         | 0.4764          |
| JLH com Mutual Information       | 0.1860         | 0.4676          |
| Chi Square                       | 0.1891         | 0.4739          |
| Google Normalized Distance (GND) | 0.1837         | 0.4609          |
| Percentage                       | 0.0167         | 0.0425          |
| Contagem absoluta (sem escore)   | <b>0.1945</b>  | <b>0.4883</b>   |

#### 4.5. Avaliação de diferentes formas de composição de resultados no RecomRI

O objetivo desse experimento é comparar a qualidade do RecomRI com uma única consulta, que é parametrizada por uma determinada quantidade de serviços visitados, com a gerada por consultas separadas para cada um dos serviços, mesclando os resultados separados para compor uma recomendação final.

**Configuração experimental** – Foi avaliada a recomendação padrão gerada com uma única consulta com as seguintes características: (i) contagem absoluta para cálculo de relevância dos serviços relacionados (melhor estratégia de acordo com os resultados da Seção 4.4); (ii) tamanho do histórico: 3; (iii) tamanho da recomendação: 10 serviços. A qualidade da recomendação padrão foi comparada com a estratégia de consultar separadamente cada um dos 3 serviços, mesclando o resultado final (vide Algoritmo 1).

**Resultados** – Os resultados comparativos são exibidos na Tabela 4. Embora ambas as estratégias apresentem valores muito próximos de precisão e revocação médias, a estratégia de consultas separadas é mais interessante do ponto de vista arquitetural, devido à sua simplicidade.

**Tabela 4: Resultados comparativos de precisão e revocação para diferentes estratégias de composição de resultados no RecomRI. Os melhores resultados estão em negrito.**

| Estratégia de Composição      | Precisão Média  | Revocação Média |
|-------------------------------|-----------------|-----------------|
| Consulta única                | 0.123064        | 0.546236        |
| Mescla de múltiplas consultas | <b>0.125447</b> | <b>0.556619</b> |

## 5. Discussão

Os resultados do experimento 1 (Seção 4.3) indicam que o cenário de recomendação com o RecomRI, implementado no Elasticsearch, apresenta qualidade superior ao ALS-100 nas condições definidas. Vale destacar que o ALS precisou ser executado em um cluster Hadoop, infraestrutura mais complexa e de manutenção mais onerosa que um cluster Elasticsearch. Portanto, o experimento foi fundamental para apoiar uma primeira decisão arquitetural quanto à tecnologia de recomendação a ser usada, indicando que a solução baseada em recuperação de informação seria capaz de oferecer bons resultados com uma arquitetura mais simples.

Os resultados do experimento 2 (Seção 4.4), por sua vez, indicam que não foi necessário utilizar métodos avançados para cálculo de relevância de itens no RecomRI, pois o método mais simples baseado em contagem absoluta já apresentou resultados competitivos. O experimento 3 mostrou que a estratégia de mesclagem dos resultados de consulta separadas para montar uma recomendação final fornece um resultado tão bom quanto o da consulta única. Ademais, tendo em vista questões de eficiência, a estratégia de consultas separadas promove o benefício de manter um cache menor, conforme discutido anteriormente no texto.

Os três experimentos demonstram que é possível alcançar resultados competitivos, em termos de qualidade de recomendação, com a abordagem de recuperação de informação (RecomRI), que é menos custosa computacionalmente do que a baseada em fatoração de matrizes. Estes resultados foram importantes para a decisão de utilizar em produção o RecomRI, com as configurações mais simples de cálculo de relevância e recomendações com mescla de múltiplas consultas.

Observando estatísticas atuais coletadas do Portal gov.br (março de 2022), foram mais de 5 milhões de cliques em links oferecidos pela recomendação no portal. Ainda nesse período, a taxa de conversão efetiva de quem chegou a solicitar o serviço recomendado estava em torno de 11%, percentual acima da taxa média global de conversão de sítios de comércio eletrônico, que alcançou 2,86% em 2020 [INVESP, 2022]. Vale ressaltar que o Portal gov.br não é um sítio de comércio eletrônico e diversas circunstâncias específicas contribuem para o interesse ou não de um cidadão por um dado serviço em dado momento. Mais importante do que seu valor absoluto é a observação dessa taxa de conversão ao longo do tempo e a análise de sua variação.

## 6. Conclusão

Este trabalho abordou a construção do sistema de recomendação do Portal gov.br – que provê ao cidadão uma melhor experiência na busca de serviços que são de seu interesse –, bem como os resultados de experimentos que guiaram as suas principais decisões arquiteturais. Destacam-se as seguintes contribuições: (1) discussão sobre as restrições de arquitetura que levaram à solução de recomendação proposta; (2) demonstração da competitividade e melhor custo-benefício da solução baseada em recuperação de informação em comparação com método tradicional baseado em fatoração de matrizes (ALS); (3) análise comparativa de métodos de cálculo de relevância de recomendação; (4) avaliação de diferentes formas de composição de resultados na solução proposta.

Trabalhos futuros incluem adaptar o sistema de recomendação do Portal gov.br para contemplar sessões de navegação de usuários autenticados, viabilizando assim que o cidadão tenha acesso à mesma recomendação de serviços, independente do canal de acesso ao Portal gov.br. Um outro trabalho deixado para o futuro é a recomendação de outros itens de conteúdo do portal, como notícias e artigos, de forma a possibilitar ao cidadão uma experiência mais rica na utilização do portal, ao receber conteúdos relevantes, não apenas serviços, ao seu perfil de navegação.

## **Agradecimentos**

Os autores agradecem ao Serpro, com destaque aos times de desenvolvimento envolvidos na solução, e à SGD/ME (Secretaria de Governo Digital do Ministério da Economia), em especial ao Departamento de Experiência do Usuário de Serviços Públicos, pela parceria na construção da solução de recomendar serviços aos usuários do Portal gov.br, bem como por permitir compartilhar o conhecimento e resultados deste projeto com a comunidade científica.

## **Referências**

- AGGARWAL, C. C. (2016) “Recommender systems”. In *Springer International Publishing*, Volume 1.
- AL-HASSAN, M., LU, H., & LU, J. (2009) “A framework for delivering personalized e-government services from a citizen-centric approach”. In *Proceedings of the 11th International Conference on Information Integration and Web-based Applications & Services*, p. 436-440.
- AYACHI, R. et al. (2016) “Proactive and reactive e-government services recommendation”. In *Universal Access in the Information Society*, 15(4), p. 681-697.
- BAEZA-YATES, R.; RIBEIRO NETO, B. (1999). “Modern information retrieval”, Volume 463.
- FERREIRA, F. et al. (2020) “Investigating Multimodal Features for Video Recommendations at Globoplay”. In *Fourteenth ACM Conference on Recommender Systems*, p. 571-572.
- GADEPALLY, V. N. et al. (2016) “Recommender systems for the department of defense and intelligence community”. In *Lincoln Laboratory Journal*, 22(1).
- GOMEZ-URIBE, C. A.; HUNT, N. (2015) “The netflix recommender system: Algorithms, business value, and innovation”. In *ACM Transactions on Management Information Systems (TMIS)*, 6(4), p. 1-19.
- GOVERNO DO BRASIL - A. (2022) “O que é o gov.br”. Disponível em: <<https://www.gov.br/sobre>>. Acesso em 19 de março de 2022.
- GOVERNO DO BRASIL - B. (2022) “Painel de monitoramento de serviços federais”. Disponível em: <<https://painelservicos.servicos.gov.br/>>. Acesso em 19 de março de 2022.

- GUTPA, U. et al. (2020) “The architectural implications of facebook's dnn-based personalized recommendation”. In *2020 IEEE International Symposium on High Performance Computer Architecture (HPCA)*, p. 488-501.
- HU, Y.; KOREN, Y.; VOLINSKY, C. (2008) “Collaborative filtering for implicit feedback datasets”. In *Eighth IEEE international conference on data mining*, p. 263-272.
- INVEP (2022) “The Average Website Conversion Rate by Industry”. Disponível em: <<https://www.invespro.com/blog/the-average-website-conversion-rate-by-industry>>. Acesso em 18 de maio de 2022.
- KOREN, Y.; BELL, R. (2015) “Advances in collaborative filtering”. In *Recommender systems handbook*, p. 77-118.
- LINDEN, G.; SMITH, B.; YORK, J. (2003) “Amazon.com recommendations: Item-to-item collaborative filtering”. In *IEEE Internet computing*, 7(1), p. 76-80.
- LOGESH, R.; SUBRAMANIASWAMY, V.; VIJAYAKUMAR, V. (2018) “A personalised travel recommender system utilising social network profile and accurate GPS data”. In *Electronic Government, an International Journal*, 14(1), p. 90-113.
- LU, J. et al. (2010) “BizSeeker: a hybrid semantic recommendation system for personalized government-to-business e-services”. In *Internet Research*.
- LU et al. (2013) “A web-based personalized business partner recommendation system using fuzzy semantic techniques”. In *Computational Intelligence*, 29(1), p. 37-69.
- MACEDO, D. et al. (2020) “Uma ferramenta para recomendação de visualização de dados governamentais abertos”. In *Anais do VIII Workshop de Computação Aplicada em Governo Eletrônico*, p. 96-107.
- MOISINHO, A. M. et al. (2021) “Modelo Conceitual para Sistemas de Recomendação voltados a Governo Eletrônico com o uso do Cadastro Base do Cidadão”. In *Anais do IX Workshop de Computação Aplicada em Governo Eletrônico*, p. 215-226.
- RESNICK, P.; VARIAN, H. R. (1997) “Recommender systems”. In *Communications of the ACM*, 40(3), p. 56-58.
- RICCI, F.; ROKACH, L.; SHAPIRA, B. (2011) “Introduction to recommender systems handbook”. In *Recommender systems handbook*, p. 1-35. Springer, Boston, MA.
- SEMERCI, O. et al. (2019) “Homepage personalization at spotify”. In *Proceedings of the 13th ACM Conference on Recommender Systems*, p. 527-527.
- SUN, N. et al. (2021) “Enhanced Collaborative Filtering for Personalized E-Government Recommendation”. In *Applied Sciences*, 11(24), p. 12119.