

MANUAL MOSTVALID

v1.0.11

SUMÁRIO

1. INTRODUÇÃO.....	1
2. PROPÓSITO.....	1
3. CONSUMO DO MOSTVALID.....	3
3.1. Authentication.....	3
3.2. Content Extraction.....	3
3.2.1. Fluxo de única requisição.....	3
3.2.2. Fluxo de múltiplas requisições.....	3
3.3. Process Status.....	4
3.4. Processo de Análise Humana.....	5
4. REGRAS DE GERAÇÃO DE TAREFAS.....	7
4.1. Regras pré-configuradas no cadastro do cliente.....	7
4.2. Regras enviadas na requisição da API.....	8
4.3. Campos depreciados.....	9
5. ORIENTAÇÕES.....	9
5.1. Uso da Ferramenta.....	9
5.2. SmartDocs Base.....	9
5.3. Política de Segurança da Informação.....	9

1. INTRODUÇÃO

A solução *mostQI* disponibiliza a tipificação e extração de documentos através da rota *Content Extraction*¹. Esta rota se baseia no mapeamento de diferentes modelos de documentos, através da construção de *SmartDocs* (processadores especializados de extração).

A natureza de alguns documentos pode implicar em uma grande variação de modelos, como as notas fiscais e, nestes casos, recomenda-se usar a rota *Content Extraction* apoiada pelo *mostValid*.

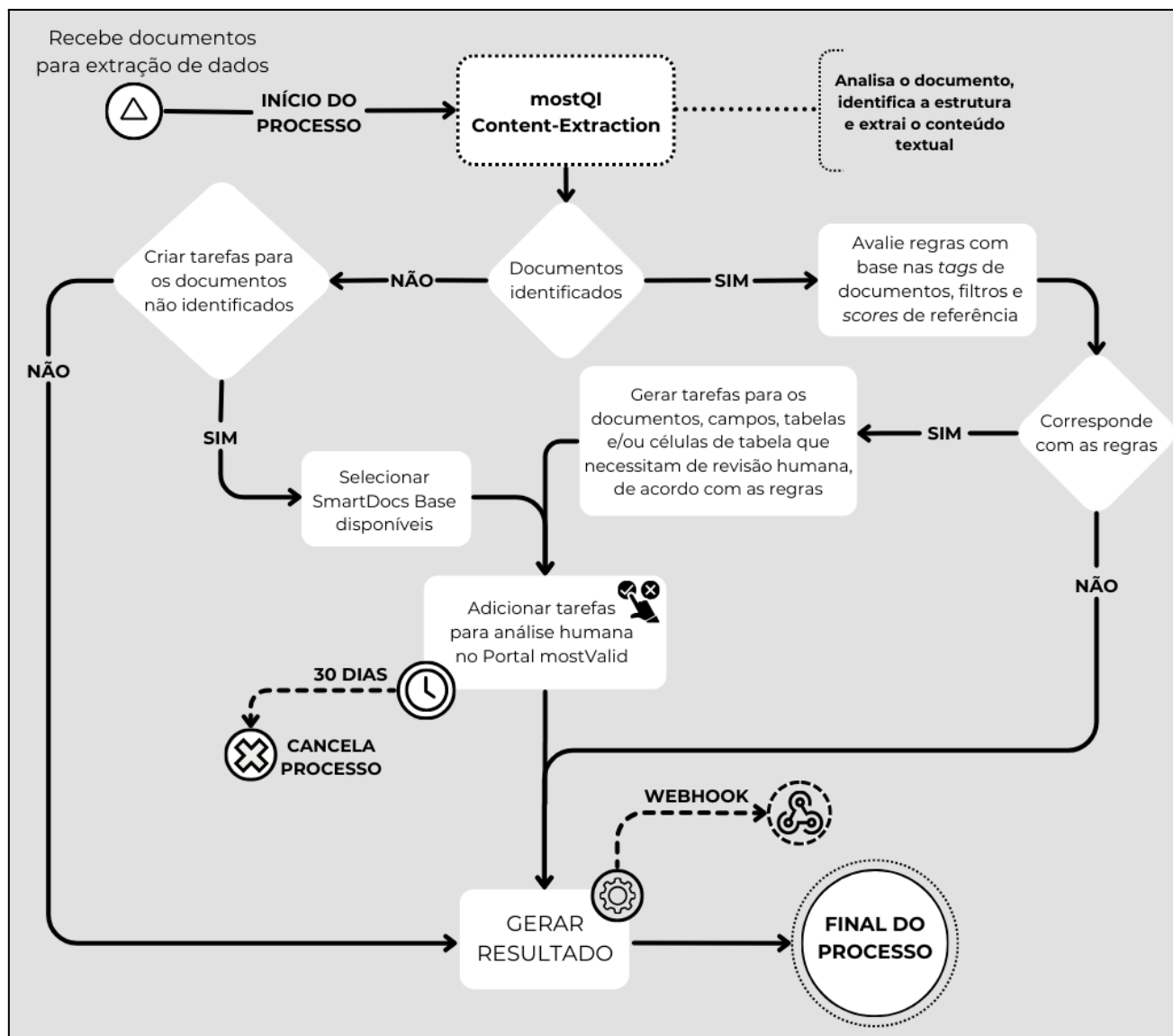
O *mostValid* apoia a operação permitindo que o operador do cliente possa trabalhar manualmente novos modelos com *SmartDocs* ainda não desenvolvidos, além de realizar validações de dados de leitura com *score* baixo. Todo o trabalho manual feito pelo operador é analisado por um time de melhoria contínua do *mostQI*, que avalia os novos modelos e, quando possível, cria *SmartDocs* para eles, visando reduzir a operação manual do nosso cliente.

O *mostValid* retém temporariamente a imagem durante seu ciclo de vida da melhoria contínua. As imagens que retornaram *scores* baixos são armazenadas para serem apresentadas nas tarefas aos operadores da empresa-cliente e posteriormente na avaliação do processo de melhoria contínua do time MOST, a fim de reduzir o índice de trabalho manual. Durante a permanência da imagem, ela não pode ser acessada de nenhuma forma que não seja pelo time de melhoria contínua e, finalizado este ciclo, em 48 horas ela é deletada.

2. PROPÓSITO

O *mostValid* é uma plataforma composta por API com métodos assíncronos para processamento, qualificação, tipificação e extração de informações a partir de imagens de documentos (utilizando o *mostQI*). Combinada a um portal eletrônico que possibilita a colaboração homem-máquina, garante uma operação fluida no processo do cliente, mesmo considerando a existência de novos documentos não treinados no *mostQI*.

¹ Acesso em: <*Content Extraction*>.



3. CONSUMO DO MOSTVALID

O processo de consumo do serviço de extração do mostValid é muito similar ao processo de extração do *mostQI* v4.

3.1. Authentication

Para cada rota de serviço a ser consumida, deve-se primeiro gerar uma *token* temporária consumindo a rota de autenticação², inserindo a *client key* cadastrada do *mostQI*.

² Acesso em: <Authentication>.

3.2. Content Extraction

Após a autenticação, o cliente pode consumir a extração de maneira similar a rota *Content Extraction*³ assíncrona do *mostQI*.

Esta rota permite dois tipos de fluxo de extração:

3.2.1. Fluxo de única requisição

O fluxo de única requisição permite o envio de arquivo em uma única tentativa, ao qual será executada a extração dos dados dos documentos presentes no arquivo e abertura da tarefa de análise humana, caso o documento se enquadre em alguma regra de validação do cliente.

3.2.2. Fluxo de múltiplas requisições

O fluxo de múltiplas requisições foi projetado para auxiliar nos processos de integração (*onboarding*), nos quais o responsável pela captura das imagens dos documentos precisa ter a flexibilidade de realizar várias tentativas (*operations*). Cada tentativa será tratada como um processo (identificado pelo *processId*) dentro de uma mesma operação (atributo *operation*), antes de abrir uma tarefa de análise humana no *mostValid*.

Finalizada as tentativas de captura da *operation*, o processo segue normalmente, retornando o resultado da extração do melhor documento (*score* mais alto) e abrindo a tarefa de análise humana, caso o melhor documento se enquadre em alguma regra de validação do cliente.

O resultado da requisição ao método *Content Extraction* será um *response* composto por um código de processo, *processId*, e um código *operationId*, quando se optar por um fluxo de múltiplas requisições.

O cliente deverá consultar periodicamente a situação do processo na rota de *status*⁴ até obter o resultado da extração, ou optar por receber os retornos em um servidor de *call back*, passando os parâmetros necessários.

³ Acesso em: <*Content Extraction*>.

⁴ Ver item 4.3. *Process Status*.

3.3. **Process Status**

Com o *processId*, o cliente pode consultar o *status* do processamento do seu arquivo enviando uma requisição na rota *Process Status*⁵. Os dados ficarão disponíveis para consulta no prazo máximo de 1 hora após a finalização do processo, ou seja, a partir do momento que a situação do processo assumir um dos *status* finalísticos: *finished*, *error* ou *anceled*.

O produto desta rota é um contrato similar ao contrato do *mostQI*. Para detalhes técnicos dos métodos acesse a documentação do *mostValid API*⁶.

Uma tarefa de validação permanecerá na fila do *mostValid* pelo período máximo de 30 dias corridos, a contar da data de sua abertura. Caso não seja realizada dentro deste prazo, o *status* da tarefa e o *processStatus* serão atualizados automaticamente para *CANCELED*.

Caso o *processStatus* apresentado seja *ERROR*, o cliente poderá entrar em contato com o suporte informando o *processId* da requisição através do e-mail suporte@most.com.br.

3.4. **Processo de Análise Humana**

O processo de análise das tarefas será iniciado após a alteração do *processStatus* para *to_analysis*, que indicará que o documento foi enviado para análise e está aguardando a avaliação humana do operador no [Portal mostValid](#).

Após o início da avaliação da tarefa, o *processStatus* é alterado para *in_analysis*. Quando há dúvidas por parte do operador, este poderá direcionar a tarefa para um supervisor, sendo um usuário com mais experiência, que irá realizar a tarefa fazendo com que o *processStatus* retorne para *to_analysis* até que um supervisor a valide.

Quando a tarefa for finalizada, o *processStatus* será alterado para *finished* e o *mostValid* trará todos os dados extraídos/revisados e a informação do *status* da revisão/modificação do escopo de cada campo e célula de tabela.

⁵ Acesso em: <*Process Status*>.

⁶ Acesso em: <Métodos da API *mostValid*>.

Os possíveis *status* relacionados ao documento, aos campos, tabelas e células de tabelas são:

- **MOSTQIIOCR**: O item não foi enviado para validação pelo revisor, pois conforme as regras do cliente, a extração feita já é considerada válida;
- **MODIFIED**: O item foi enviado para validação, pois foi verificado que ele necessitava de revisão, conforme as regras definidas pelo cliente, e o seu conteúdo original foi modificado pelo revisor;
- **UNREADABLE**: O item foi enviado para validação, pois foi verificado que ele necessitava de revisão conforme as regras definidas pelo cliente, sendo marcado como ilegível pelo revisor;
- **REVIEWED**: O item foi enviado para validação, pois foi verificado que ele necessitava de revisão, conforme as regras definidas pelo cliente, e o seu conteúdo original não foi modificado pelo revisor;
- **DEPRECATED**: Trata-se de um item com nome do campo depreciado e que não será exibido no Portal mostValid para análise durante a tarefa de revisão. Recomenda-se considerar os nomes de campos correspondentes atualizados. Confira a listagem atualizada de documentos e respectivos campos [aqui](#).

Comportamento na comparação de tabelas:

- Caso a tabela não seja visualizada pelo Revisor devido às configurações do cliente, todas as células, cabeçalho e a tabela serão atualizados como **MOSTQIIOCR**;
- Caso algum cabeçalho tenha sido alterado pelo revisor, ou alguma coluna tenha sido adicionada na tabela, todas as células, cabeçalhos e a tabela ficarão com o valor salvo durante a revisão e seu *status* será **MODIFIED**, visto que se trata de uma nova tabela.
- Caso a tabela tenha sido visualizada pelo revisor, devido às configurações do cliente, será executado uma rotina que verificará se o revisor alterou a célula:

- Caso tenha alterado a célula, esta será marcada como *MODIFIED*;
- Caso não tenha alterado a célula, esta será marcada como *REVIEWED*, independente de ter entrado na regra do usuário ou não;
- A tabela será marcada como *REVIEWED* caso nenhuma célula tenha sido alterada, e será marcada como *MODIFIED* caso pelo menos uma célula seja alterada pelo usuário.
- Caso o revisor tenha visualizado a tabela e adicionado mais linhas: Seja N o número de linhas da tabela original e M o número de linhas da tabela alterada pelo revisor, as N linhas serão comparadas com as M primeiras linhas alteradas. A comparação pode não ser feita com a mesma linha da tabela original, visto que o revisor pode ter adicionado uma linha no lugar de uma linha antiga. As linhas que sobrarem (M - N) serão marcadas como *MODIFIED*, visto que serão consideradas novas linhas. A tabela também será marcada como *MODIFIED*.
- Caso a tabela tenha sido visualizada pelo revisor e ele tenha deletado uma ou mais linhas da tabela: Seja N o número de linhas da tabela original e M o número de linhas da tabela alterada pelo revisor, as M linhas serão comparadas com as N primeiras linhas da tabela original. Esta comparação pode não ser feita com as linhas correspondentes na tabela original dependendo da posição na qual o revisor deletou a linha. Logo, a tabela será marcada como *MODIFIED*.

4. REGRAS DE GERAÇÃO DE TAREFAS

4.1. Regras pré-configuradas no cadastro do cliente

O cliente pode definir regras para a geração de tarefas de acordo com suas necessidades, com o apoio do time de *Customer Success*. Existem duas categorias de regras: gerar tarefas para documentos não identificados e gerar tarefas para documentos identificados.

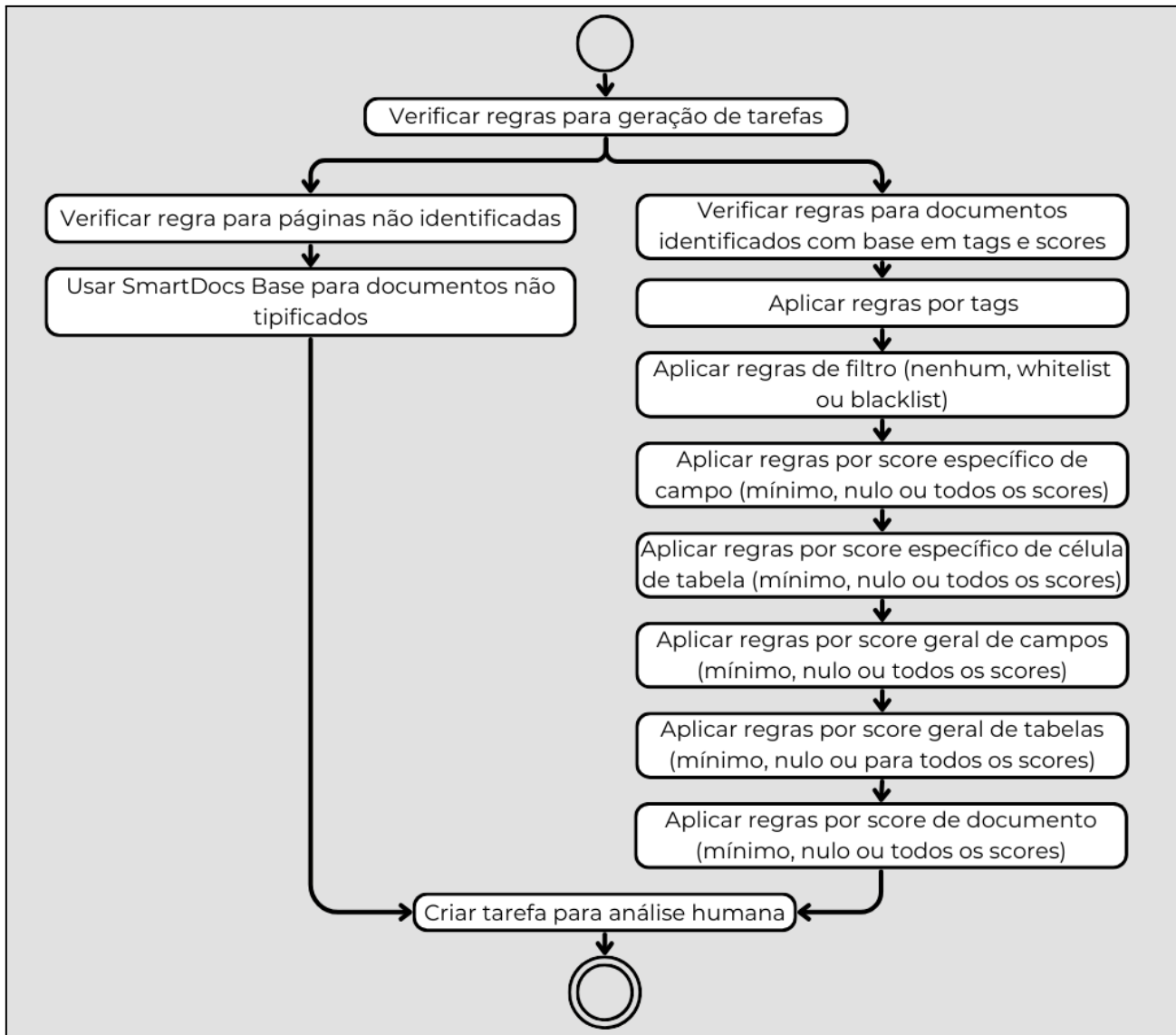
Para os documentos não identificados, é possível utilizar os *SmartDocs* Base para a seleção prévia de campos. Os *SmartDocs* Base permitem a classificação humana e o preenchimento de informações durante a análise no Portal *mostValid*.

Para os documentos identificados, as regras para geração de tarefas podem ser configuradas de várias maneiras:

- Filtros:
 - *Whitelist*: Ao estabelecer uma *whitelist* de campos e/ou tabelas, as configurações das regras de *score* serão aplicadas exclusivamente aos campos especificados nela;
 - *Blacklist*: As configurações das regras de *score* se aplicarão a todos os campos e/ou tabelas, exceto àqueles definidos na *blacklist*;
 - Sem filtro: As configurações se aplicarão a todo o conteúdo extraído do documento.
- *Score* mínimo específico por campo ou célula de tabela;
- *Score* geral mínimo por documento, campo e/ou tabela.

Tarefas serão criadas se o *score* resultante da extração de um documento, campo ou célula de tabela for um valor abaixo do valor mínimo definido na regra (o valor de referência pode variar entre 0 e 1). Também é possível gerar tarefas independentemente dos *scores* (para todos os campos) ou para *scores* nulos.

Caso exista filtro de *whitelist*, as regras de *score* serão aplicadas exclusivamente aos itens filtrados e caso exista filtro de *blacklist*, o campo ou célula de tabela não será avaliado, independentemente do *score*.



4.2. Regras enviadas na requisição da API

É possível informar as regras para geração de tarefas, diretamente na requisição ao mostValid, através do parâmetro *generationRules*. Este parâmetro aceita uma lista de no máximo 5 regras e irá sobrepor as regras configuradas no portal (caso existam) para o processamento do documento enviado na requisição. Veja detalhes sobre o objeto *generationRules* na documentação do mostValid API⁷.

⁷ Acesso em: <Métodos da API mostValid>.

4.3. Campos depreciados

É importante destacar que campos depreciados não serão exibidos no mostValid, sendo recomendado substituí-los pelos campos atualizados correspondentes.

Confira a listagem atualizada de documentos e respectivos campos [aqui](#).

5. ORIENTAÇÕES

5.1. Uso da Ferramenta

O dinamismo da ferramenta e o trabalho realizado em documentos enviados para revisão é exibido em um [vídeo demonstrativo](#). Para informações sobre o portal do mostValid acesse o [Manual de Uso](#).

5.2. SmartDocs Base

A validação da extração de dados de um documento a partir do uso de um SmartDoc Base é exibido em um [vídeo demonstrativo](#).

5.3. Política de Segurança da Informação

As normas de segurança da informação e prevenção contra incidentes estão contidas na [Política de Segurança da Informação](#).