

**Tema: Implantação de Boas Práticas de Gestão e Melhoria da Qualidade dos Serviços Prestados pela Receita Federal do Brasil.**

# Menção Honrosa

---

**Batimento Automatizado de Documentos na Importação - BatDoc.**

**Autor:**

**Diego de Borba Barbosa**

## ANEXO 3

Portaria RFB nº 1207, de 29/07/2016

## RESUMO DA MONOGRAFIA

Nº: 016

### INSCRIÇÃO

**TÍTULO DO TRABALHO:**

**Batimento Automatizado de Documentos na Importação - BatDoc**

### A) OBJETIVOS BÁSICOS;

Este trabalho apresenta métodos de análise de dados não estruturados os quais foram aplicados em uma solução que confere a exatidão e a correspondência das informações prestadas na declaração de importação (DI) em relação àquelas constantes dos documentos instrutivos. Assim, visando a inovação e a agilidade deste processo de trabalho, apresenta-se uma ferramenta de batimentos automatizados de documentos na Importação (BatDoc) .

Os objetivos desta ferramenta são: otimizar a análise documental nos despachos de importação permitindo o direcionamento do tempo para análises complexas relativas ao despacho; auxílio para que erros/inconsistências não passem despercebidos; gerar uma interface de relatório simples e intuitiva com links web ágeis; permitir flexibilidade de configuração visando uma maior efetividade da solução.

### B) METODOLOGIA UTILIZADA;

De forma descritiva, introduzimos a aplicabilidade do BatDoc na análise do despacho de importação. Em seguida, como o BatDoc foi pensado, sua definição, arquitetura, métodos de análise de dados construídos, estratégia de integração com outros *softwares*, interface de acesso, interface do relatório e resultados. Nas demonstrações do relatório BatDoc foram utilizadas DI reais resguardando o sigilo fiscal das informações. Ainda, em trabalhos derivados aborda-se o potencial dos métodos de análise construídos no contexto do gerenciamento de risco e revisão aduaneira. Por último, aborda-se trabalhos futuros em relação a ferramenta BatDoc e uma conclusão.

## **C) ADEQUAÇÃO DO TRABALHO AOS CRITÉRIOS DE JULGAMENTO:**

### **I – CRIATIVIDADE E INOVAÇÃO;**

O BatDoc representa uma proposta inovadora para área aduaneira da RFB, mais especificamente no contexto da análise dos documentos instrutivos no despacho de importação. Funciona como uma ferramenta de suporte à análise documental na importação. Esse suporte ocorre por meio do relatório BatDoc que acusa as divergências encontradas no confronto DI versus documentos. Tal solução demonstra a potencial aplicação dentro da RFB dos métodos de análise de dados não estruturados.

### **II – RELAÇÃO CUSTOS VERSUS BENEFÍCIOS;**

O desenvolvimento da solução não gerou despesa adicional à RFB visto que foram utilizados *softwares* internos ou cuja licença a RFB já possui. Os benefícios são:

#### **1 - Otimizar a análise documental nos despachos de importação:**

Na sistemática anterior, os documentos em papel eram - em regra - organizados sobre uma mesa e os dois monitores tinham a função de visualização da DI e dos sistemas, ou seja, no método impresso há uma agilidade de conferência intrínseca.

No atual cenário de recepção automática dos documentos instrutivos digitalizados, a conferência estritamente documental é realizada - em regra - com a visualização dos documentos e DI em dois monitores. Logo, ao entregar um relatório que acusa as divergências encontradas, a análise dos documentos instrutivos fica facilitada na medida em que se faz a conferência manual das inconsistências apontadas pelo relatório.

#### **2 – Eficácia na análise documental (Erros formais não apurados):**

Tal relatório, ainda, auxilia para que erros não passem despercebidos já que toda informação não localizada é apontada no relatório para fins de análise manual.

#### **3 – Interface de acesso:**

O BatDoc é visualizado em uma das abas internas do Aniita com uma interface de relatório simples, intuitiva, centralizada e com a possibilidade de visualização do relatório em navegador web para melhor utilização dos links web. Assim, a geração do relatório torna-se algo natural para o usuário habituado com esse sistema.

#### **4 - Possíveis ganhos não mensuráveis:**

A demonstração da aplicabilidade de análise de dados não estruturados, obtidos a partir de arquivos digitalizados, destaca para a instituição os potenciais benefícios destes métodos de

análise. O desenvolvimento do BatDoc busca instigar o surgimento de mais projetos desse tipo em outras áreas da Receita Federal, ou até mesmo na Administração Pública.

### **III - AUMENTO DE PRODUTIVIDADE;**

A otimização gerada pelo BatDoc por meio da entrega de um relatório que permite conhecer os acertos e erros encontrados na conferência documental permite a emissão da primeira intimação/interrupção no Siscomex ou desembaraço num menor espaço de tempo. Logo, é possível redirecionar o tempo para análises mais complexas relativas ao despacho, por exemplo, análise de classificação fiscal e emissão de relatórios de outros sistemas;

Não foi possível mensurar quantitativamente a ferramenta, mas a avaliação macro da equipe de despacho aduaneiro do autor foi de incentivar a solução pois valoriza o trabalho na equipe de despacho.

### **IV - VIABILIDADE DE IMPLEMENTAÇÃO;**

A disponibilização do BatDoc ocorrerá em basicamente 3 fases como componente dentro do Aniita. Logo, todos os usuário que já utilizam o Aniita farão uso do BatDoc de forma natural. O BatDoc já está em utilização e aprimoramento na unidade do autor deste trabalho o que demonstra que é totalmente viável(fase 1). O próximo passo (fase 2) será a utilização por 7 unidades-piloto para os modais aéreo, marítimo, ferroviário e rodoviário. O estágio do piloto será uma ótima oportunidade de crescimento da ferramenta, tendo em vista que estará em constante utilização. A expectativa é que os *feedbacks* gerem melhorias nas especificidades de cada modal contribuindo para gerar um relatório o mais efetivo possível. Após a fase do piloto e da devida avaliação da Coana espera-se a disponibilização nacional(fase 3).

### **V - MELHORIA DA QUALIDADE DOS SERVIÇOS PRESTADOS E DOS RESULTADOS ESTRATÉGICOS;**

O BatDoc possui relação com os seguintes Objetivos do Mapa Estratégico da RFB, a seguir relacionados:

#### **Perspectiva de Resultados:**

**- Garantir a arrecadação necessária ao Estado, com eficiência e aprimoramento do sistema tributário:**

Ainda que em âmbito aduaneiro a arrecadação não seja o objetivo principal, o BatDoc auxilia na verificação de conformidade das declarações de importação, atuando na identificação de informações errôneas ou incompletas. Estes erros ou omissões são punidos pelas normas

aduaneiras com multas, por exemplo, a multa de 1% do valor aduaneiro por prestação de informação errada, incompleta ou imprecisa na DI, Art. 711 do Regulamento Aduaneiro (Decreto 6.759/2009).

**- Garantir a segurança e agilidade no fluxo internacional de bens, mercadorias e viajantes:**

Na área aduaneira o controle aduaneiro é o objetivo principal, logo, minimizar o tempo da análise documental até a primeira intimação/interrupção do Siscomex ou desembaraço torna o despacho aduaneiro mais célere.

**Perspectiva de Processos:**

- Ampliar a aplicação da análise de riscos nos controles tributários e aduaneiros
- Incentivar o cumprimento das obrigações tributárias e aduaneiras
- Contribuir para a facilitação do comércio internacional e do fluxo de viajantes, em articulação com os demais órgãos

**Perspectiva de Gestão e Suporte:**

- Desenvolver competências, valorizar pessoas e adequar o quadro de pessoal às necessidades institucionais
- Viabilizar recursos e otimizar sua aplicação para suprir as necessidades de infraestrutura e tecnologia
- Assegurar um modelo organizacional que favoreça a integração e a inovação nos processos

Na Cadeia de Valor da RFB o(s) processo(s) de trabalho a que se refere a monografia são:

- ✓ Controle Aduaneiro
- ✓ Governança de Tecnologia da Informação

**D) PRINCIPAIS LIÇÕES APRENDIDAS.**

O trabalho em equipe demonstrou-se imprescindível tendo em vista que a cada *feedback* foi possível melhorar a eficiência da ferramenta. O aprendizado técnico foi excelente, o sucesso da solução será um sucesso dos membros da equipe que direta ou indiretamente participaram da construção da solução. Ainda, comprovou-se a viabilidade de utilização de OCR para análise de dados não estruturados dentro da RFB com consequentes benefícios e sem custos extras.



## **15º Prêmio de Criatividade e Inovação da RFB**

### **Implantação de Boas Práticas de Gestão e Melhoria dos Processos de Trabalho e dos Serviços Prestados ao Contribuinte**

#### **Batimento Automatizado de Documentos na Importação (BatDoc)**

**2016**



## Sumário

1 INTRODUÇÃO.....	3
2 APLICABILIDADE NO DESPACHO DE IMPORTAÇÃO.....	6
2.1 Etapas do Despacho.....	6
2.2 Recepção Automática de Documentos na Importação.....	7
2.3 Exame Documental.....	7
2.4 Aplicabilidade do BatDoc.....	8
3 BATDOC.....	9
3.1 Definição e Objetivos.....	9
3.2 Dados Estruturados.....	11
3.3 Dados Não Estruturados.....	12
3.4 Expressões Regulares.....	12
3.5 OCR.....	13
3.6 OCR Abbyy FineReader.....	13
3.7 Arquitetura do BatDoc.....	14
3.8 Pré-Mapeamento dos Documentos.....	18
3.9 Método de Análise 1 - Dado Estruturado (DI) → Dado Não Estruturado (Documentos Instrutivos).....	20
3.10 Método de Análise 2 - Dado Não Estruturado (Documentos Instrutivos) → Dado Estruturado (DI).....	25
3.11 Interface de Acesso.....	30
3.12 Resultados e Interface do Relatório.....	31
3.12.1 Navegabilidade Geral.....	32
3.12.2 Navegabilidade em Adição.....	37
3.12.3 Navegabilidade em Itens da Adição.....	37
3.12.4 Alerta Vermelho.....	40
3.12.5 Links Web Automáticos.....	43
4 TRABALHOS DERIVADOS.....	46
4.1 BatDoc.....	46
4.1.1 Aplicabilidade no Gerenciamento de Risco.....	46
4.1.2 Aplicabilidade em Revisões Aduaneiras de DI.....	47
4.2 Potencial dos Métodos de Análises.....	49
5 TRABALHOS FUTUROS.....	49
6 CONCLUSÃO.....	49
REFERÊNCIAS:.....	51
ANEXO 1.....	52



## **Batimento Automatizado de Documento na Importação (BatDoc)**

### **Resumo**

Este trabalho apresenta métodos de análise de dados não estruturados, os quais foram aplicados em uma solução que confere a exatidão e a correspondência das informações prestadas na declaração de importação (DI) em relação àquelas constantes nos documentos instrutivos. Assim, visando à inovação e à agilidade desse processo de trabalho, apresenta-se uma ferramenta de batimento automatizado de documentos na Importação (BatDoc).

### **1 INTRODUÇÃO**

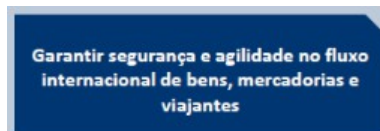
A Secretaria da Receita Federal do Brasil (RFB) exerce a administração tributária federal e a aduaneira em nosso país. Em âmbito aduaneiro tem a função precípua de promover o controle aduaneiro sob veículos, mercadorias e viajantes de forma a cumprir a legislação aduaneira vigente.

O Estado não pode dispensar esse controle por maior que seja seu nível de livre-cambismo. Se por um lado esse controle serve para assegurar a correta arrecadação dos tributos incidentes na importação, principalmente, assegurando-se a aplicação do Princípio do Tratamento Nacional, por outro lado, tal controle transcende o interesse arrecadatório. O foco extrafiscal é precípua, visto que é indispensável que o Estado controle as mercadorias importadas e exportadas visando à segurança e à saúde pública, à defesa comercial, aos bens sensíveis,



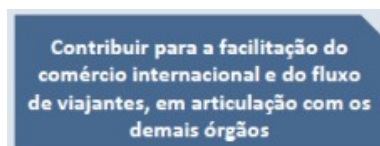


dentre outros interesses relevantes para a sociedade. Tal controle deve garantir não só a segurança, mas também a agilidade no fluxo internacional, conforme consta nos objetivos de resultado do Mapa Estratégico RFB 2016-2019:



*Fonte: RFB*

Conforme mencionado, o controle aduaneiro é essencial para a sociedade e para o Estado, porém, com a crescente globalização, a comunidade internacional também busca promover a agilidade e facilitação do comércio internacional. Essa facilitação consta como objetivos de processo no Mapa Estratégico RFB 2016-2019:



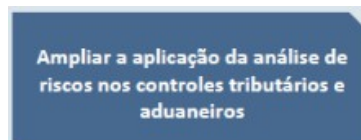
*Fonte: RFB*

O tema “Facilitação de Comércio Internacional” não é novo, mas é atual. Temos, por exemplo, o Artigo VIII do GATT 47 que já apresentava às partes contratantes diretrizes de simplificação nos controles aduaneiros. Pode-se citar exemplos mais recentes como o Programa de Operador Econômico Autorizado (OEA) em 2015 e a Ratificação do Acordo de Facilitação do Comércio da Organização Mundial do Comércio (OMC) em 2016, entre outros.

É notório que há uma aparente dualidade entre o controle aduaneiro e a facilitação do comércio. Para simplificar e agilizar os processos aduaneiros são imprescindíveis, como contrapartida, por exemplo, simplificações administrativas e tributárias e/ou investimentos em soluções que permitam de fato a agilidade nas



análises tributárias e aduaneiras, bem como um intenso fortalecimento dos gerenciamentos de riscos. Esse fortalecimento do gerenciamento de risco também consta como objetivos de processo no Mapa Estratégico RFB 2016-2019:



*Fonte: RFB*

No campo de inovações, cita-se o Analisador Inteligente e Integrado de Transações Aduaneiras (Aniita) (COUTINHO, 2012), o Sistema de Seleção Aduaneira por Aprendizado de Máquina (SISAM) (JAMBEIRO FILHO, 2015) e a Anexação Digital de Documentos (VICOMEX, 2015) como importantes e imprescindíveis soluções para o atingimento dos objetivos de agilidade na liberação de mercadorias. Essas relevantes ferramentas são utilizadas para gerenciar riscos aduaneiros e também na análise de despachos de importação, as quais propiciam à RFB uma maior eficácia e efetividade no fluxo de trabalho.

De forma alinhada com a busca por agilidade, simplificação e facilitação de comércio, apresenta-se o estudo de dois métodos de análise de dados não estruturados com uma abordagem no despacho de importação. Desenvolveu-se uma solução capaz de realizar o batimento automatizado de documentos (BatDoc).

Essa ferramenta permite verificar, de forma automatizada, a correspondência das informações prestadas na DI em relação aos documentos instrutivos do despacho, entregando, como produto, um relatório para a análise do Auditor-Fiscal responsável pelo procedimento fiscal. O BatDoc foi desenvolvido e integrado ao

Aniita. Ele utiliza outros softwares que são *open source* ou com licenças já adquiridas pela RFB, logo, uma excelente relação custo-benefício para a RFB.

Neste trabalho apresenta-se esse *software*, o BatDoc, e - ainda que de maneira ampla - como os métodos de análise de dados não estruturados podem ser explorados em outras áreas. No próximo tópico apresenta-se detalhadamente a etapa do despacho de importação que se pretende agilizar com o BatDoc.

## 2 APLICABILIDADE NO DESPACHO DE IMPORTAÇÃO

### 2.1 Etapas do Despacho

Para se ter precisão em relação a etapa do despacho de importação que se pretende otimizar por meio dos confrontos automatizados de informações, destacamos abaixo o já conhecido fluxograma, Figura 1, com as etapas do despacho de importação.

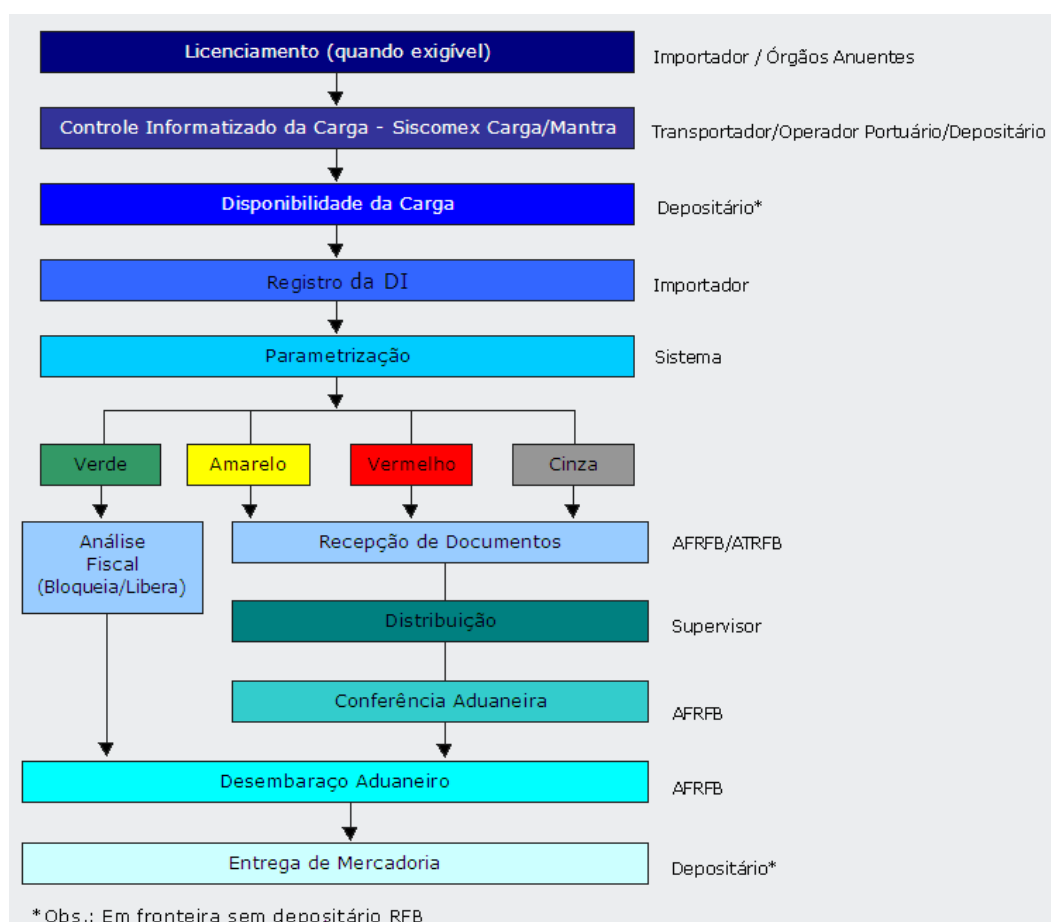


Figura 1: Etapas do despacho; Fonte:RFB



Observa-se que os canais de parametrização amarelo, vermelho e cinza exigem a recepção dos documentos instrutivos e a consequente conferência aduaneira, exame documental e conferência física (conforme o canal). De acordo com o que dispõe a legislação vigente, o exame documental é atividade comum a todos esses canais.

Assim, o BatDoc tem como objeto a automatização de parte das atividades realizadas no exame documental. No próximo tópico, apresenta-se a sistemática da recepção automática dos documentos instrutivos.

## **2.2 Recepção Automática de Documentos na Importação**

Desde 1º de julho de 2015 os documentos instrutivos da DI, e demais requerimentos apresentados no curso do despacho, são apresentados à RFB em meio digital, conforme o art. 19 da IN SRF Nº 680/2006.

A recepção dos documentos instrutivos em meio digital é realizada pelo módulo Anexação de Documentos Digitalizados(VICOMEX, 2015), acessível por meio do Portal Único de Comércio Exterior. Tal módulo permite ao importador, e seus representantes legais perante o Siscomex, apresentar aos órgãos intervenientes no comércio exterior documentos de forma digitalizada. Assim, após vincular o dossiê digital à DI ocorre a recepção automática dos documentos perante a RFB.

No próximo tópico demonstra-se o enquadramento normativo do exame documental e se especifica as atividades automatizadas no exame documental.



### **2.3 Exame Documental**

O exame documental é um procedimento fiscal destinado a realização de muitas verificações de conformidade administrativas e tributárias. Dentre as verificações do exame documental, o Auditor-Fiscal verifica a exatidão e correspondência das informações prestadas na declaração em relação àquelas constantes nos documentos que a instruem (uma espécie de *checklist* dos documentos instrutivos versus a DI), conforme Art. 25, inciso II, IN SRF Nº 680/2006.

É notório que a aferição da correta prestação das informações na DI seja realizada ordinariamente pelos documentos instrutivos obrigatórios, pois são estes que retratam toda a negociação comercial internacional e devem conter as informações prestadas na DI.

Essa é a etapa ou estritamente a verificação que buscamos simplificar e agilizar utilizando-se o BatDoc. Resumidamente, o BatDoc analisa a correspondência das informações prestadas na DI em relação aos documentos instrutivos e emite um relatório para subsidiar a análise do Auditor-Fiscal.

### **2.4 Aplicabilidade do BatDoc**

No atual cenário de recepção automática dos documentos instrutivos digitalizados, a conferência estritamente documental é realizada - em regra - com a visualização dos documentos e DI em dois monitores. Na sistemática anterior, os documentos em papel eram organizados sobre uma mesa e os monitores tinham a função de visualização da DI e dos sistemas, ou seja, no método impresso há uma

agilidade de conferência intrínseca.

Assim, nasceu a ideia inovadora de uma solução que fosse capaz de conferir a exatidão e correspondência das informações prestadas na DI em relação às constantes dos documentos instrutivos, na forma de uma ferramenta de suporte para a análise documental.

### 3 BATDOC

#### 3.1 Definição e Objetivos

O BatDoc é uma solução que realiza o *download* de todos os documentos do Anexação Eletrônica e utiliza um *OCR* para realizar batimentos automatizados com campos pré-mapeados da DI, entregando um relatório que acusa as divergências encontradas. Trata-se de um batimento documental automatizado dos documentos instrutivos em relação à DI. A ferramenta permite conhecer os acertos e erros encontrados gerando agilidade na análise documental. Tal agilidade externaliza-se com o desembaraço ou primeira intimação num menor espaço de tempo.

De maneira macro pode-se representar o BatDoc, Figura 2, com entradas dos documentos digitalizados e das informações prestadas na DI, muito processamento, inclusive com um *OCR*, e na saída um relatório *.html* que é um formato web.

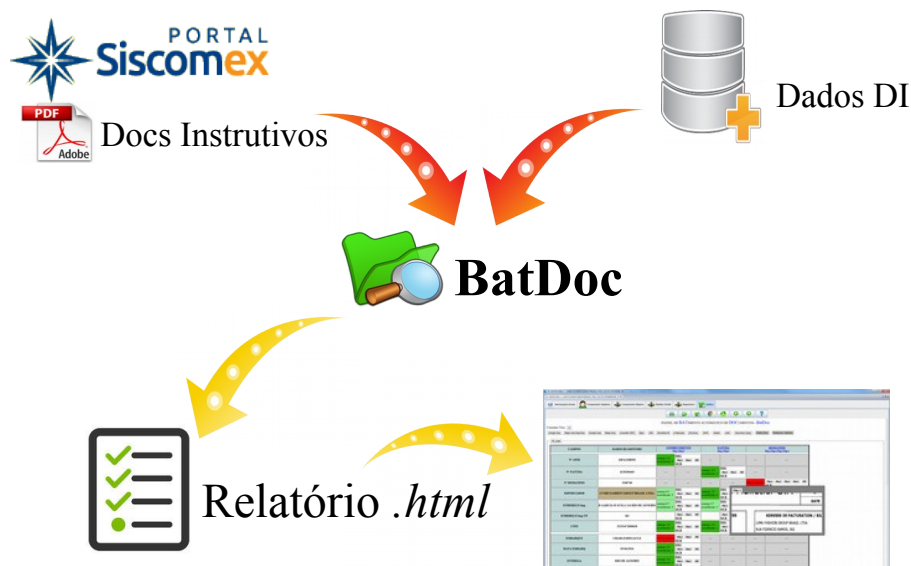


Figura 2: BatDoc; Fonte: Elaborado pelo autor.



Os objetivos desta ferramenta são os seguintes: otimizar a análise documental nos despachos de importação, permitindo redirecionar o tempo para análises mais complexas; auxiliar para que erros/inconsistências não passem despercebidos; gerar uma interface de relatório simples e intuitiva com *links web* ágeis; permitir a inclusão de novos padrões de busca, novas equivalências de dados, novos *links* e alteração das regras oriundas do mapeamento inicial de documentos, visando à adaptação a peculiaridades de cada unidade e a uma maior efetividade da solução;

O BatDoc apresenta importante auxílio quando o Auditor-Fiscal responsável pela DI limita a conferência aduaneira ao motivo da seleção, ou seja, quando faz uso da faculdade prevista no § 1º do Art. 24 da IN SRF 680/2006. Nessa situação, o relatório gerado pelo BatDoc, por permitir verificar inconsistências no exame documental como um todo e de forma ágil, proporciona mais segurança para que o Auditor-Fiscal faça uso dessa faculdade normativa.

Ainda que em âmbito aduaneiro a arrecadação não seja o objetivo principal, o BatDoc auxilia na verificação de conformidade das declarações de importação, atuando na identificação de informações errôneas ou incompletas. Estes erros ou omissões são punidos pelas normas aduaneiras com multas, por exemplo, a multa



de 1% do valor aduaneiro por prestação de informação errada, incompleta ou imprecisa na DI, Art. 711 do Regulamento Aduaneiro (Decreto 6.759/2009).

A contribuição imediata do BatDoc insere-se na análise de DI's em canal vermelho e amarelo e nas retificações de DI em canal verde a pedido do importador. Uma potencial contribuição futura é a utilização do BatDoc para análise de DI's em canal verde com aplicação no contexto do gerenciamento de risco e de revisões aduaneiras. Essa utilização potencial será abordada no tópico “Trabalhos Derivados”.

A disponibilização do BatDoc ocorrerá em basicamente 3 fases como componente dentro do Aniita. Logo, todos os usuário que já utilizam o Aniita farão uso do BatDoc de forma natural. O BatDoc já está em utilização e aprimoramentos na unidade do autor deste trabalho o que demonstra que é totalmente viável(fase 1). O próximo passo (fase 2) será a utilização por 7 unidades-piloto para os modais aéreo, marítimo, ferroviário e rodoviário. O estágio do piloto será uma ótima oportunidade de crescimento da ferramenta, tendo em vista que estará em constante utilização. A expectativa é que os *feedbacks* gerem novas ideias de melhorias contribuindo para gerar um relatório o mais efetivo possível. Após a fase do piloto e da devida avaliação da Coana, espera-se a disponibilização nacional (fase 3).

### **3.2 Dados Estruturados**

Dados estruturados são informações armazenadas em bancos de dados. Em regra, são organizados em tabelas relacionais e seus registros possuem atributos rígidos. Por exemplo, uma coluna data armazena informações dessa natureza, ou





uma coluna de valor não admite texto. Abstraindo, imaginemos uma planilha excel, com muitas tabelas. Cada tabela tem um nome, as colunas são os atributos, e cada linha é um registro de informações do mundo real. Em tal forma de organização, há uma previsibilidade do tipo de informação que se pretende extrair tendo em vista um esquema de dados pré-definido, diferentemente de dados não estruturados (TESSAROLO, 2015). Os dados da DI são armazenados e recuperados a partir de banco de dados mantido e administrado pelo SERPRO, logo, pode-se definir que tais informações são do tipo dados estruturados.

### **3.3 Dados Não Estruturados**

Basicamente são os dados que não possuem uma estrutura definida, a estrutura desses dados é particular de cada arquivo. Exemplo, arquivos de documentos textos, imagens e vídeos. (TESSAROLO).

Conforme já citado, atualmente a recepção dos documentos da DI pela RFB se dá de forma automática com a vinculação de um dossiê aos documentos instrutivos digitalizados. Tais arquivos digitalizados são dados não estruturados, em regra, imagens geradas a partir de um equipamento *scanner* do importador, exportador e/ou de seus representantes.

Logo, para ser possível a comparação de dados da DI com documentos digitalizados(arquivos de imagem), logicamente é imprescindível que essas informações fossem transformadas em um conjunto de dados alfanuméricos. Para tanto, passou-se a pesquisar um software capaz de reconhecer, processar e extrair texto de imagens. Aborda-se em seguida o conceito de *Regex*, o *OCR* utilizado para



extração de texto(dado não estruturado), assim como a arquitetura planejada.

### **3.4 Expressões Regulares**

Expressões regulares ou do inglês *regular expression* (REGEX) são formas concisas de identificar padrões de cadeia de caracteres. São muito utilizadas em mecanismos de *software* para validação de entrada de informações em sistemas, ou seja, seu principal objetivo é validar formatos de dados na entrada de programas. Assim, por exemplo, é possível validar um e-mail, um cpf ou uma senha de sistema conforme políticas específicas do programa.

Utilizou-se expressões regulares para identificação de padrões particulares de caracteres nos Métodos de Análise a serem apresentados em tópico próprio.

### **3.5 OCR**

A tecnologia *Optical Character Recognition* (OCR), reconhecimento ótico de caracteres, possibilita a identificação e extração de alfanuméricos a partir de imagens ou mapa de bits. Essa tecnologia é muito utilizada, por exemplo, em radares rodoviários e na distribuição interna de correspondências nos Correios. Neste trabalho, a utilização de tal tecnologia é pré-requisito para o alcance dos objetivos visto a necessidade de conversão dos arquivos digitalizados(imagens) para arquivos de texto(pesquisáveis).



### 3.6 OCR Abbyy FineReader

Os arquivos que são submetidos ao *OCR* são arquivos digitalizados, em regra os documentos instrutivos digitalizados estão no formato *Portable Document Format* (PDF) tendo em vista a possibilidade de paginação. Como o foco não é o estudo e implementação da tecnologia *OCR*, realizou-se pesquisas, testes de viabilidade e de custo-benefício, primeiramente com uma ferramenta *open source* disponibilizada pela Google, o *OCR Tesseract*, e uma segunda tentativa com o *OCR Abbyy FineReader*.

Realizamos alguns testes com o *Tesseract*, porém identificamos que tal *OCR* tem um grau de assertividade mediano quando submetido ao processamento de imagens digitalizadas. Por exemplo, imagens que o olho humano consegue ler sem grandes dificuldades, tal *OCR* obteve resultados muito prejudicados. Logo, concluiu-se que o custo-benefício de tal *OCR* era negativo visto que emitir um relatório, que é o produto final, que não localiza informações nada agregaria. Por ser um projeto de código aberto está em constante desenvolvimento, assim pode ser objeto de mais testes no futuro (TESSERACT, 2006).

O *software Abbyy FineReader 9.0.1* é um *OCR* cuja licença já foi adquirida pela RFB, logo, supera-se a questão orçamentária dos custos. Nos testes realizados com tal *OCR* o grau de assertividade foi excelente, quando comparado com o *Tesseract*, para os mesmos documentos digitalizados. A versatilidade na exportação dos arquivos processados é um outro ponto positivo.

Utilizou-se o módulo *Abbyy HotFolder* o qual permite criar uma rotina de monitoramento de pastas de acordo com uma tarefa *Abbyy* configurada. Essa tarefa

é um arquivo do tipo *.hft*, uma estrutura de dados que guarda informações para execução da tarefa, por exemplo: pasta de monitoramento(*input*), pasta de saída(*output*), processamento completo ou rápido, formato de arquivo de saída e Cpf do usuário(ABBYY, 2016).

### 3.7 Arquitetura do BatDoc

O BatDoc foi desenvolvido no ambiente *IDE Eclipse*, na linguagem de programação *java* e utilizando controlador de versões *Git*.

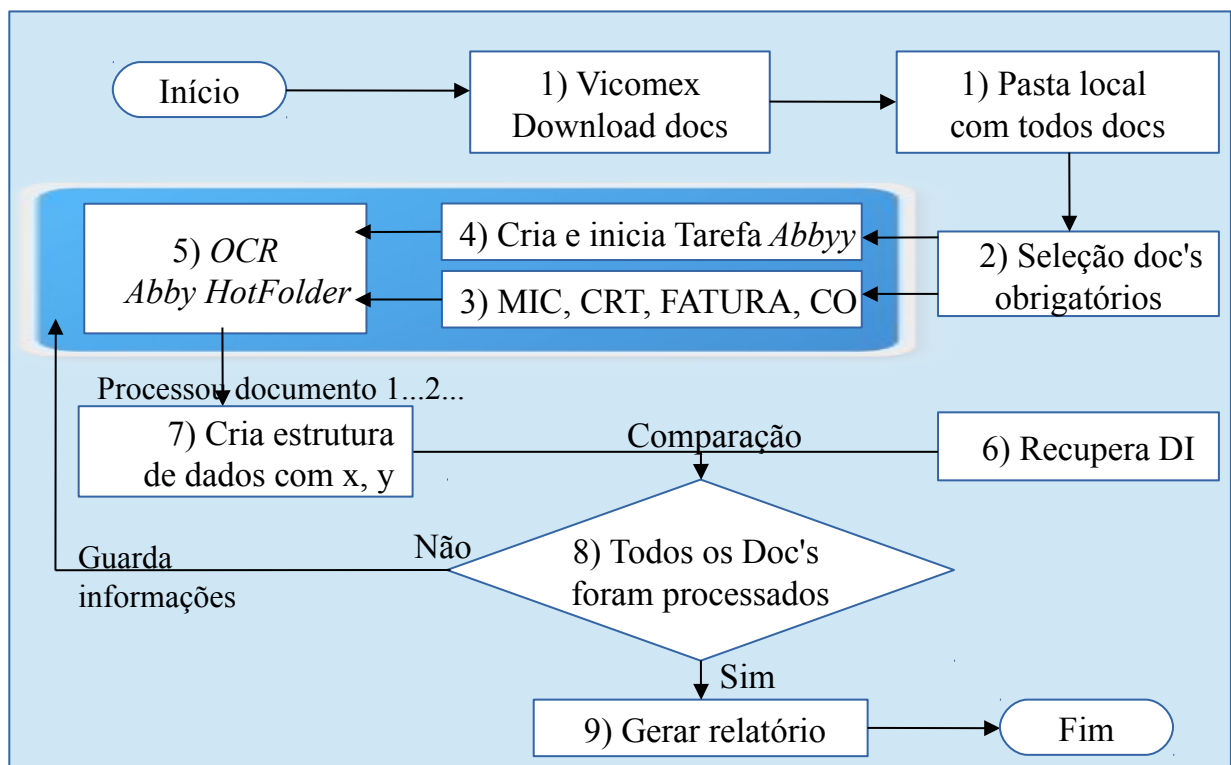


Figura 3: Arquitetura BatDoc; Fonte: Elaborado pelo autor.

A estratégia e fluxo dos algoritmos do BatDoc para aplicar a análise de dados não estruturados no contexto do despacho aduaneiro de importação pode ser



descrita de forma macro conforme a Figura 3:

1) realizar o *download* automatizado de todos os documentos digitalizados do sistema Anexação Digital para a pasta local de trabalho do Auditor-Fiscal; 2) selecionar aqueles que são documentos instrutivos obrigatórios, dentre todos os arquivos; 3) copiar os arquivos selecionados para uma pasta de monitoramento do *OCR-Abbyy*; 4) criar um arquivo de tarefa do tipo específico do *OCR* com configurações pré-determinadas; 5) executar a citada tarefa para que o *OCR* inicie o monitoramento da pasta onde estão as cópias dos arquivos selecionados, realizando a extração do texto das imagens; 6) recuperar os dados da DI do Aniita; 7) após o processamento pelo *OCR* de cada documento, criar uma estrutura de dados com coordenadas x e y para fins de recortes de informações localizadas; 8) iniciar uma rotina de comparações, especificadas conforme mapeamento, entre a estrutura de dados de cada documento e os dados da DI; efetuar recortes no PDF das informações localizadas; guardar os resultados das comparações e aguardar o término do processamento do próximo documento; 9) após o processamento de todos os documentos, gerar um relatório, no formato *.html*, com as informações localizadas, não localizadas, alertas e *links* úteis.

Assim, o BatDoc, realiza o *download*, arquivos digitalizados de determinada DI, para a pasta do usuário que foi configurada(item 1 da arquitetura) e faz a seleção dos documentos obrigatórios (item 2 da arquitetura). Essa seleção é necessária tendo em vista que o importador pode anexar documentos complementares, por exemplo, requerimentos diversos, documentação probatória de interesse de outros órgãos anuentes, dentre outros. Tendo selecionados os

documentos, o BatDoc copia os arquivos para a pasta de *input* do OCR que foi configurada na tarefa *Abbyy* (item 3 da arquitetura).

Para que não ocorra falha nessa etapa, ou seja, documentos obrigatórios que fiquem fora do processamento, é indispensável que o interveniente tenha anexado ao dossiê digital o tipo de arquivo correto, por exemplo, “fatura comercial” e não utilize tipos residuais como “outros documentos”.

Abaixo observe-se a Figura 4 que demonstra a perspectiva de funcionamento do *OCR-Abbyy*:



Figura 4: Fluxo OCR Abbyy HotFolder; Fonte: Abbyy

O BatDoc cria e inicializa a tarefa *Abbyy* (item 4 da arquitetura), após isso o processamento do *OCR* fica em segundo plano na visão do usuário. Dessa forma o *OCR* inicia o monitoramento da pasta *input*, e sempre que houver arquivos digitalizados novos (item 1 da Figura 4) tenta processá-los (item 2 da Figura 4) e



exporta o resultado, PDF de texto(pesquisável), para pasta *output* configurada(item 3 da Figura 4) (item 5 da arquitetura).

Configurou-se o arquivo de saída em *PDF* pesquisável, do tipo texto sob imagem, assim não se perde as características do documento original.

A rotina do BatDoc que estava monitorando esta pasta *output* verifica que determinado documento foi processado e utiliza esse *PDF* pesquisável para criar uma estrutura de dados não estruturados com coordenadas x e y (item 7 da arquitetura) para cada um dos documentos processados. Por exemplo, a Figura 5 demonstra uma imagem do campo 15 do conhecimento de embarque rodoviário(CRT):

15 Gastos a pagar Gastos a pagar	Monto Remitente Valor remetente	Moneda Moeda	Monto Destinatário Valor destinatário
<b>Frete / Frete</b>			
	<b>3350.00</b>	<b>USD</b>	<b>.00</b>
<b>Seguro / Seguro</b>			
	<b>.00</b>	<b>USD</b>	<b>.00</b>
<b>Otros / Outros</b>			

Figura 5: Campo 15 CRT; Fonte: Adaptado pelo autor.

Abaixo uma exemplificação da estrutura de dados gerada pelo BatDoc:

[119,471]

3350.00

Explicação:



- 3350.00 : Conjunto alfanumérico cujas coordenadas foram descritas, ou seja, o valor do frete que consta no CRT;

- [119,471]: Coordenadas do primeiro alfanumérico do conjunto “3”; 119 é a coordenada x e 471 é a coordenada y para localização no PDF;

O BatDoc utiliza a estrutura de dados não estruturados com coordenadas x/y para iniciar a rotina de batimentos automatizados e para efetuar os recortes das informações localizadas em cada documento (item 8 da arquitetura). A rotina de batimentos automatizados será melhor apresentada nos tópicos de Métodos de Análise de Dados, bem como o relatório *.html* será abordado em tópico próprio.

### **3.8 Pré-Mapeamento dos Documentos**

Para confrontar os dados estruturados constantes na DI com os dados não estruturados oriundos de cada documento instrutivo, método de análise abordado no item 3.9 desse trabalho, se faz necessário saber quais informações prestadas na DI são originalmente prestadas em quais documentos. Os documentos instrutivos da DI divergem também dependendo do modal, por exemplo, o conhecimento de embarque rodoviário possui algumas informações diferentes do conhecimento de embarque marítimo. Logo, esse mapeamento dos documentos nos diferentes modais, nos possibilita saber quais informações prestadas na DI devem ser objeto de busca em cada tipo de documento.

Realizou-se um mapeamento, no qual concluímos que as informações a serem verificadas nos documentos instrutivos são essencialmente as contidas no





anexo I deste trabalho.

Uma das telas de configurações do BatDoc, Figura 6, possibilita ao Auditor-Fiscal alterar o mapeamento de batimentos, ou seja, há uma flexibilidade para que seja possível o atendimento de eventuais peculiaridades das unidades. Assim, por exemplo, é possível adicionar a comparação do número da fatura comercial no conhecimento de embarque rodoviário(CRT) como um padrão no Método de Análise DI → Docs bastando selecionar o *checkbox* referente ao campo versus documento respectivos, e salvar a configuração. Ainda, sempre será possível restaurar a configuração original do mapeamento dos documentos.

Em relação aos batimentos automatizados/comparações(item 8 da arquitetura), desmembrou-se tal rotina em dois métodos de análise os quais foram construídos e estão em constante aprimoramento.



Configurar Métodos de Análise - BatDoc

Método Análise DI->Docs

Campos	CO Mercosul	CRT	BL-HBL	AWB	Fatura	MIC-DTA	TIF-DTA	Romaneio
<b>Capa DI</b>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
N Certificado	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
N CRT/BL-HBL/AWB	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
N Fatura	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
N MIC-DTA/TIF-DTA	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
N Romaneio	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Importador	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Endereço Imp	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
CNPJ Imp	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Adquirente	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Endereço Adq	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
CNPJ Adq	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Transportador	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Embarcação	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Cidade Embarque	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Data Embarque	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Qtde Volumes	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Cidade Entrega	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Peso Bruto	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Peso Líquido	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
VMLE	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Dicionário de Dados      Web Links      Ajustar Flexibilidade

Método Análise Docs->DI

CNPJ       CPF       NOME EMPRESAS       NCM

Incluir Novos Padrões      Restaurar Config      Salvar      Sair

Figura 6: Tela Configurar Métodos de Análise; Fonte: Elaborado pelo autor.

### 3.9 Método de Análise 1 - Dado Estruturado (DI) → Dado Não Estruturado (Documentos Instrutivos)

Esse método foi pensado para tornar possível a comparação de dados da DI perante os documentos instrutivos. Logo, seguindo-se a estratégia inicial, foi necessário realizar o mapeamento dos documentos instrutivos, ou seja, quais informações constantes na DI deveriam ser localizadas em cada documento. Após isso, foi necessário a criação de algoritmos baseados em expressões regulares para realizar o confronto “dados da DI → documentos”. Importante frisar que esse método parte da DI para os documentos (Método de Análise DI → Docs).

A melhor forma de explicar a construção desse método de análise é a

exemplificação. Pois bem, na Figura 7 temos uma informação da DI - ENDEREÇO Exp\_1, endereço do exportador da adição 1 - a ser localizado nos diversos documentos instrutivos. O BatDoc monta um conjunto de expressões regulares por palavras dessa informação da DI. Nesse caso, temos um endereço com 7 palavras, logo, 7 padrões de procura. Por exemplo, em relação a fatura comercial a localização foi parcial visto que foram localizadas 5 palavras das 7 totais.

Adição_1		ORIGEM Pág.1		CONHECIMENTO Pág.1		FATURA Pág.1		MANIFESTO Pág.1   Pág.2		ROMANEIO Pág.1	
CAMPOS	DADOS DI:1614638740										
EXPORTADOR_1	[REDACTED]	certeza 3/5 ocorrências: 1	IMG Pág.1 pdf OCR	Não localizado	IMG Pág.1 pdf OCR	certeza 3/5 ocorrências: 2	IMG Pág.1 pdf OCR	Não localizado	IMG Pág.1   Pág.2 pdf OCR	Não localizado	IMG Pág.1 pdf OCR
ENDEREÇO Exp_1	AV. A. FLEMING 002238 MUNRO BUENOS AIRES	certeza 6/7 ocorrências: 1	IMG Pág.1 pdf OCR	certeza 6/7 ocorrências: 1	IMG Pág.1 pdf OCR	certeza 5/7 ocorrências: 3	IMG Pág.1 pdf OCR	Pág: 1 [REDACTED]			
NCM_1	20041000	certeza: 1/1 ocorrências: 1	IMG Pág.1 pdf OCR	---	---	---	---				
INCOTERM_1	FCA	---	---	certeza: 1/1 ocorrências: 1	IMG Pág.1 pdf OCR	certeza: 1/1 ocorrências: 1	IMG Pág.1 pdf OCR				

Figura 7: Tela relatório BatDoc - Método Análise DI -> Docs; Fonte: Elaborado pelo autor.

Essa procura é realizada por blocos alfanuméricos dos dados não estruturados que foram criados para cada documento. Utiliza-se recursos de *Regex* para desconsiderar acentuação, pontuação, caracter especial, maiúscula/minúscula, quando for o caso. Assim, é possível chegar em um mesmo nível de comparação de dados. Por exemplo: na DI foi informado o CNPJ 11222333000122, já nos documentos tal informação pode vir em muitos formatos, por exemplo: 112223330001-22, 11222333/000122, 11.222.333/000122, 11.222.333/0001-22, dentre outros. O mesmo acontece com nomes de empresas e endereços.

Ao localizar tal informação, armazena-se basicamente a linha da estrutura de dados, a quantidade de padrões localizados, e as coordenadas x e y. Essas

coordenadas são utilizadas para fazer um recorte no *PDF* da posição da localização para fins de conferência manual no relatório BatDoc. Tanto a linha como o respectivo recorte da informação localizada são exibidos no relatório como fundamentos da localização positiva.

A certeza foi definida com o padrão de no mínimo 60% de palavras localizadas. Logo, uma localização 3/7 iria ser apontada no relatório como não localizada. A interface do relatório será objeto de estudo em tópico próprio. Ainda, construiu-se um menu para ajustar a flexibilidade de localização de alguns campos da DI, Figura 8:

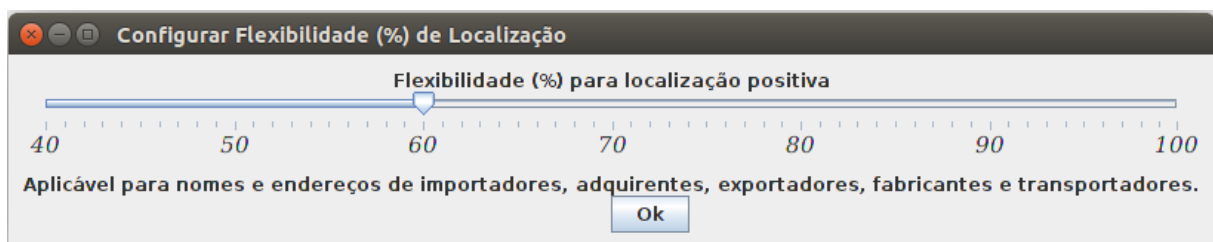






Figura 8: Tela Configurar Flexibilidade - Método Análise DI -> Docs; Fonte: Elaborado pelo autor.

A partir da experiência adquirida com algoritmos anteriores mais rígidos, concluímos que essa abordagem era necessária para possibilitar flexibilidade na localização de campos com predominância de texto. No exemplo acima da Figura 7, a omissão/abreviatura/erro de carácter/erro de leitura do *OCR* - que é comum em documentos impressos - não impediria que o BatDoc localizasse a informação. Importante frisar que essa flexibilidade foi definida apenas para campos com predominância de texto, ou seja, para os nomes e endereços de importadores, adquirentes, exportadores, fabricantes e transportadores declarados na DI. Em tais

campos a localização é sempre para um conjunto total maior que 1, por exemplo: 3/5 (localizadas 3 de 5 palavras); 6/6 (localizadas 6 de 6 palavras).

Essa mesma abordagem confere rigidez para campos numéricos visto que sempre formam a expressão regular com 1 padrão. Assim, definimos que a flexibilidade configurada anteriormente não se aplica para campos numéricos. Por exemplo a Figura 9, observa-se um peso líquido de 25075,00Kg o padrão é apenas o número 25075,00 ou 25075. Logo, a localização positiva exige certeza 1/1, ou seja, localização integral da informação numérica.

PESO BRUTO Kg	25130	---	---	certeza: 1/1 ocorrências: 1	 Pág. 1 Pdf OCR
PESO LÍQUIDO Kg	25075	---	---	certeza: 1/1 ocorrências: 1	 Pág. 1 Pdf OCR
VOLUMES (BIG BAG)	25	---	---		 Pág. 1 Pdf OCR
VMLE US\$	2119.84	---	---	certeza: 1/1 ocorrências: 1	 Pág. 1 Pdf OCR

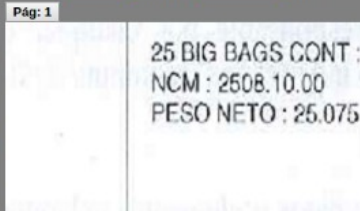


Figura 9: Tela relatório BatDoc - Método Análise DI -> Docs; Fonte: Elaborado pelo autor.

Ainda, com o intuito de reforçar a localização de informações criou-se um dicionário de dados, ou seja, expressões que são consideradas equivalentes para fins de análise documental. Por exemplo, av = avenida, ltda = limitada, imp = importadora, com = comércio, cia=companhia, dentre tantas outras.

Embora não seja pré-requisito para o funcionamento da ferramenta, tendo em vista a flexibilidade descrita anteriormente, vislumbra-se um aumento considerável na eficiência dos batimentos automatizados ao considerar como correta tais informações. A partir da experiência adquirida em análise de despachos aduaneiros,

verificou-se que essa equivalência de informações é realizada naturalmente num exame documental pelo Auditor-Fiscal. Logo, não há qualquer prejuízo para a análise. Abaixo, apresenta-se a tela de configuração do dicionário de equivalências, Figura 10:



Figura 10: Tela Incluir Novas Equivalências; Fonte: Elaborado pelo autor.

Além das equivalências mais comuns já incluídas, cada Auditor-Fiscal tem a liberdade de incluir novas equivalências ou excluir as já existentes. Tal recurso foi pensado para aumentar naturalmente a capacidade de localização da ferramenta na medida em que permite a inclusão de peculiaridades das unidades em função do modal, região e línguas.

Outra decorrência dos testes com o BatDoc no Método de Análise DI → Docs, foi a percepção de que informações muito curtas podem comprometer a confiabilidade da localização positiva. Por exemplo, como localizar um peso bruto de 20Kg. Nessa situação, certamente uma data seria indicada como localização



positiva em razão do ano 2016. Logo, definiu-se que informações da DI iguais ou menores que 2 caracteres não serão objeto de busca, evitando uma localização falso positiva. Já as informações com mais de 2 caracteres serão objeto de busca com um número máximo de ocorrências esperadas para aquele documento. Quando se atinge esse número máximo de ocorrências, um aviso aparece ao lado da localização positiva para indicar a conferência manual daquela informação da DI.

O presente método de análise gera um relatório com as localizações positivas e negativas, e em nível de capa, de adição e de itens da DI.

No contexto do Método de Análise DI → Docs, não se pode garantir que tais documentos instrutivos não contenham dados além da DI, por exemplo, um nome ou Cnpj de terceira empresa não declarada. Isso ocorre porque traçar o fluxo de comparação no sentido dos dados não estruturados para a DI, exige reconhecimento de padrões viável, porém para a maioria das informações não há possibilidade de definir um padrão de reconhecimento confiável. Visando suprir tal deficiência, foi construído um outro método de análise que será abordado a seguir.

### **3.10 Método de Análise 2 - Dado Não Estruturado (Documentos Instrutivos) → Dado Estruturado (DI)**

O Método de análise dado não estruturado (documentos instrutivos) → dado estruturado (DI) visa completar a lacuna do método anterior, ou seja, suprir a impossibilidade daquele método de verificar informações relevantes nos documentos e que supostamente deveriam constar na DI mas não foram devidamente declaradas.



Por meio de *Regex*, construímos expressões para reconhecimento de padrões particulares de cadeia de caracteres nos dados não estruturados dos documentos da DI. As expressões regulares implementadas buscam identificar CNPJ, CPF, códigos da nomenclatura comum do Mercosul(NCM) e a maior parte dos nomes de empresas brasileiras. Abaixo exemplifica-se uma expressão regular capaz de reconhecer uma cadeia de caracteres que pode ser um CNPJ :

```
[^0-9][0-9]{2}[-\.]?[0-9]{3}[-\.]?[0-9]{3}[-\.\V\\]?[0-9]{4}[-\.\V\\]?[0-9]{2}[^0-9]
```

Explicação:

- `[^0-9]`: Faixa de caracteres diferente de 0 a 9;

Tal padrão foi utilizado no início e no fim da expressão para prevenir erros de localização de conjuntos de números que não são Cnpj, por exemplo, no meio do código de barras de faturas comerciais;

- `[0-9]{2}` Faixa de caracteres: 0 a 9, quantidade: 2 caracteres;

- `[-\.]?` Um traço ou ponto, opcional. Foi usado `\` no ponto, pois ele sozinho é caractere especial;

- `[0-9]{3}` Faixa de caracteres: 0 a 9, quantidade: 3 caracteres;

- `[-\.]?` Um traço ou ponto, opcional. Foi usado `\` no ponto, pois ele sozinho é caractere especial;

- `[0-9]{3}` Faixa de caracteres: 0 a 9, quantidade: 3 caracteres;

- `[-\.\V\\]?` Um traço, ou ponto, ou barra, ou contrabarra, facultativo. Foi usado `\` para evitar caractere especial;

- `[0-9]{2}` Faixa de caracteres: 0 a 9, quantidade: 2 caracteres;





Esta solução valida, por exemplo, os seguintes formatos: 11222333000133, 11.222.333/0001-33, 112223330001-33, 11222333/0001-33 11222333/000133, 11-222-333-0001-33, 11.222.333\000133, 11.222.333\0001.33, dentre outros.

Em se tratando de nomes de empresas brasileiras, para a razão social e a denominação consegue-se padronizar buscas por meio das designações especiais, conforme Direito Comercial. Por exemplo, sociedades limitadas (LTDA.), sociedade em comandita por ações (C/A), sociedade anonima (S/A), sociedade em nome coletivo (N/C), sociedade em comandita simples (C/S), cooperativa, microempreendedor individual (MEI), empresa individual de responsabilidade limitada (EIRELI), microempresa (ME), empresa de pequeno porte (EPP). As empresas individuais não seriam reconhecidas nesse tipo de busca tendo em vista a formação da firma individual sem uma designação específica.

Nesse cenário, moldou-se um algoritmo capaz de reconhecer possíveis CNPJ, CPF, NCM e nome de empresas nos documentos instrutivos. Tendo em vista que se conhece o nome do importador, adquirente, transportador, fabricante, exportador estrangeiro, CNPJ, CPF e NCM declarados na DI, pode-se verificar se tais nomes guardam compatibilidade com os padrões localizados nos dados não estruturados. Assim, sempre que ocorra esta localização nos dados não estruturados (documentos instrutivos), o BatDoc verifica a compatibilidade desse dado com os dados declarados na DI. Caso seja incompatível com os dados constantes da DI, o relatório apresenta essa incompatibilidade como um alerta vermelho na parte inferior do relatório BatDoc.

A Figura 6 demonstra a tela de configuração em que é possível desligar



algumas das funções do Método de Análise Docs → DI. Por exemplo, caso o Auditor-Fiscal não tenha interesse em alertas de CPF que não constem na DI, basta desmarcar o *checkbox* correspondente e salvar a configuração.

Desenvolveu-se uma forma de inclusão de novos padrões de busca nos documentos os quais denominamos de parâmetros, além das 4 já fixadas no programa (CNPJ, CPF, NCM, nomes empresas). Esses novos padrões podem ser simplesmente um CNPJ específico a ser monitorado ou uma nova expressão regular. Abaixo segue a Figura 11 onde é feita tal inclusão:

Identificação(nome)	Parâmetro	Observações
empresa suspeita	ABC imp e exp ltda	Abandonou mercadorias sem responder Intimaçõ...
cnpj monitorado	11.222.333/0001-44	Cnpj de Adquirente não declarado. Apurado na DI...
NCM	[^0-9][0-9]{4}[\.][0-9]{2}[-\.]?[0-9]{2}[^0-9]	Exemplo de inclusão expressão regular

Obs.: A Identificação e Parâmetro são obrigatórios para salvar o registro.  
Obs.: Ao Salvar os dados incluídos, novos campos em branco são criados para novas inserções.

OK

Figura 11: Tela Incluir Novos Padrões(Parâmetros); Fonte: Elaborado pelo autor.

Logo, qualquer informação alfanumérica pode ser incluída para fins de monitoramento/busca nos documentos instrutivos. Por exemplo, uma razão social específica, um endereço, nome de uma empresa específica, um CNPJ, enfim qualquer informação. Caso seja localizado algum padrão incluído nessa tela, ele sempre constará no relatório como um alerta vermelho.

A identificação é o nome do parâmetro que será indicado no alerta, o parâmetro é a informação alfanumérica ou expressão regular efetivamente




procurada nos documentos instrutivos(dados não estruturados), e observações são informações adicionais a serem apresentadas dentro do alerta quando incidir seu parâmetro. Todos os novos padrões de localização incluídos podem ser salvos.


Tal Método de Análise Docs → DI, possibilita uma cultura de monitoramento de documentos de forma eficiente e independente do tempo. Isso instiga e possibilita a formação de uma base de conhecimento pessoal sobre especificidades que podem vir a ser equivocadamente esquecidas no curso do despacho. Por exemplo, o Auditor-Fiscal na análise de uma DI identifica certos cuidados para análise da classificação fiscal de determinada NCM ou destaque de LI. Nesse caso, poderia incluir um novo padrão sendo o parâmetro a NCM daquela mercadoria e uma observação. Caso após alguns meses outro despacho com a mesma NCM venha a ser analisada pelo mesmo Auditor, o relatório já indicaria a incidência de um parâmetro, indicaria um alerta vermelho e possibilitaria a imediata leitura da respectiva observação/lembrete.

No estágio atual, os novos padrões de localização incluídos são salvos apenas no computador local do Auditor-Fiscal. Outra possibilidade a se pensar, seria a formação e compartilhamento dessa base de conhecimento em nível local, regional ou nacional. Nessa visão, o BatDoc indicaria alertas vermelhos para todos parâmetros que incidissem com as devidas observações.

Inicialmente pensou-se em 2 tipos de saída para o Método de Análise Docs → DI como alertas no relatório:

	alerta	trata-se de Cnpj, Ncm, Cpf ou nome de empresa cujo padrão
---	--------	---



	verde	reconhecido em documentos tenham correlação com a DI;
	alerta vermelho	trata-se de Cnpj, Ncm, Cpf ou nome de empresa, porém nesse caso que não constam na DI; e todos os novos padrões incluídos cujo parâmetro foi localizado em alguns dos documentos;

Porém, definiu-se que apenas as informações do alerta vermelho agregam valor no relatório BatDoc visto que o alerta verde é redundante, pois tais informações já foram objeto de localização no método DI → Docs.

Por exemplo, tal método de análise busca nos documentos todos os conjuntos numéricos cujo padrão enquadrar-se como possível CNPJ, verifica se tal conjunto tem correspondência com algum CNPJ informado na DI, em caso negativo constará no alerta vermelho como um indício.

### 3.11 Interface de Acesso

A interface de acesso ao BatDoc foi construída dentro do Analisador Inteligente e Integrado de Transações Aduaneiras (Aniita) (COUTINHO, 2012), que já é usado para redirecionamento de DIs e apoio ao despacho aduaneiro e que apresenta, em um único local, informações oriundas de diferentes sistemas da RFB.

*“O Aniita é um software que visa fortalecer o gerenciamento de risco aduaneiro na Receita Federal do Brasil (RFB). Os objetivos da ferramenta são: otimizar o tempo da análise de risco aduaneiro; possuir uma interface amigável; descomplicar o acesso à informação, independentemente da aptidão de cada*



*indivíduo com a tecnologia de informação (TI); e permitir a parametrização local de Declarações de Importação (DI).*” (COUTINHO, 2012).

Após o Aniita recuperar todas as informações de diversos sistemas e aplicar as parametrizações, o *download* do dossiê digital e a aplicação do BatDoc serão iniciados em segundo plano, liberando a navegabilidade no Aniita. Logo, há possibilidade de análise imediata da DI enquanto o dossiê e o relatório BatDoc não são concluídos.

Enquanto não ocorrer tal conclusão, o andamento pode ser acompanhado pela nova aba BatDoc dentro da DI no Aniita (Aniita >> DI >> Aba BatDoc). A Figura 12 demonstra a aba BatDoc com um relatório finalizado.

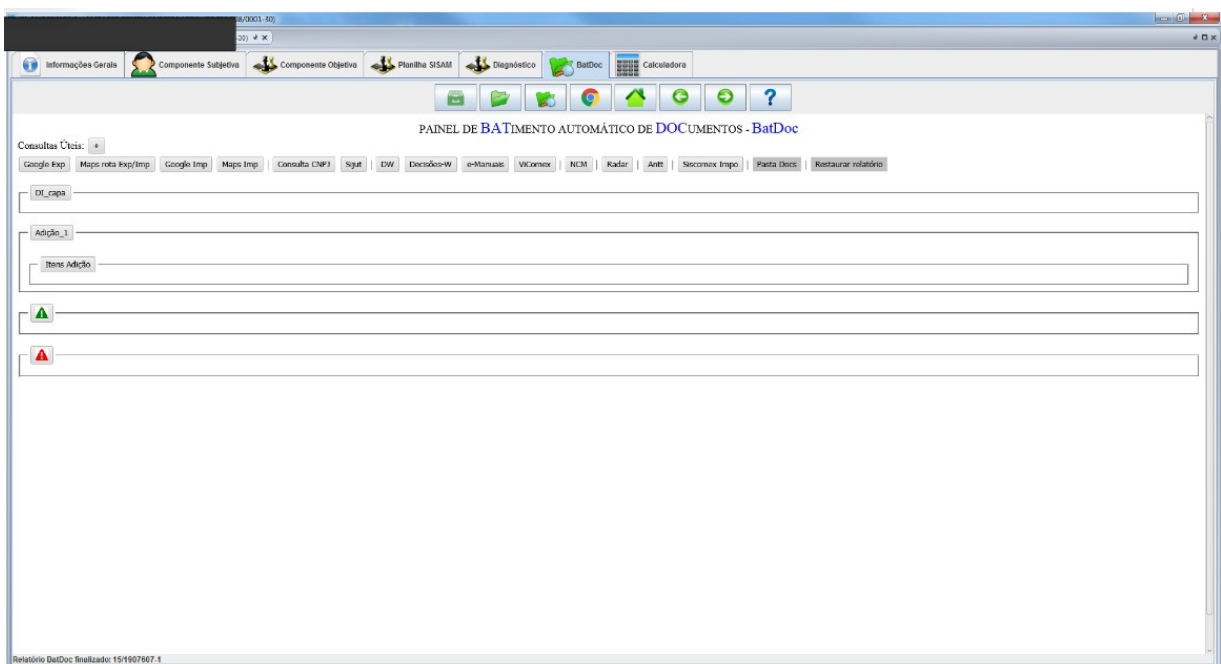


Figura 12: Interface do Aniita e aba BatDoc. Fonte: Aniita e autor

### 3.12 Resultados e Interface do Relatório



O BatDoc está previsto para começar numa fase de piloto visando


adaptações e aprimoramentos. Obtendo êxito nessa fase, as coordenações responsáveis têm entendido pela liberação em âmbito nacional da solução. Hoje o BatDoc está em uso na unidade do autor, porém não foi possível efetivar avaliações quantitativas precisas.

Em relação ao relatório, inicialmente pensou-se que os resultados do confronto de dados pudessem ser abstraídos para o usuário por meio de um relatório positivo de batimentos, ou seja, emissão de um relatório com todas as informações localizadas. Outra possibilidade seria um relatório negativo, um relatório das informações da DI cuja correspondência nos documentos instrutivos não foi localizada. Por fim, decidiu-se pela construção de um relatório misto.

Tal relatório apresenta as informações localizadas e não localizadas nos documentos instrutivos, apresenta a certeza e o número de ocorrências da localização, recortes aproximados de imagens das informações localizadas, links para os documentos instrutivos integrais, links úteis para web, e o alerta vermelho.




Abaixo descreve-se os principais botões dispostos na aba BatDoc do Aniita:

	<p>Botão <i>Download</i></p>	<p>Botão para tentar novamente o <i>download</i> do dossiê quando: ocorrer indisponibilidade temporária do <i>download</i> ou quando o <i>checkbox</i> de recuperações for desmarcado no Aniita.</p>
	<p>Botão Aplicar BatDoc</p>	<p>Botão para executar a função aplicar BatDoc quando: ocorrer indisponibilidade temporária do <i>download</i> ou do software <i>OCR</i>, ou simplesmente reaplicar os batimentos em função do acréscimo de algum novo padrão de procura.</p>

	Botão Abrir no Navegador	Para visualização do relatório diretamente no navegador e melhor utilização dos web links, recomenda-se a visualização pelo Google Chrome
---	--------------------------------	---

### 3.12.1 Navegabilidade Geral

Os métodos de análise do BatDoc demonstram basicamente a certeza e ocorrências das informação que se pretende localizar em cada documento. Segue uma legenda das cores utilizadas:

	Totalmente localizado. Ex.: Certeza 4/4; 1/1;
	Parcialmente localizado. Ex.: Certeza 3/4; 5/7;
	Não localizado. Ex.: Certeza 2/4; Certeza 3/7;

A certeza relaciona-se com a quantidade de palavras (alfanuméricos) do campo da DI que foram localizadas. Ocorrência é a quantidade de vezes que determinada informação (conjunto alfanumérico) da DI foi localizada no documento.

Por exemplo, no relatório BatDoc Figura 13, coluna ORIGEM(Certificado de Origem), no campo/linha Nº CERTIFICADO, a análise foi que se localizou totalmente o número do certificado (1 conjunto alfanumérico de um total de 1 = Certeza 1/1) uma vez no documento (ocorrências = 1). Passando o mouse sobre IMG pode-se visualizar o recorte aproximado do documento onde tal informação foi localizada.



PAINEL DE BATIMENTO AUTOMÁTICO DE DOCUMENTOS - BatDoc

Consultas Úteis: +  
 Google Exp | Maps rota Exp/Imp | Google Imp | Maps Imp | Consulta CNPJ | Sijut | DW | Decisões-W | e-Manuais | ViComex | NCM | Radar | Antt | Siscomex Impo | Pasta Docs | Restaurar relatório

CAMPOS	DADOS DI:1612008668	ORIGEM Pág.1 Pág.2	CONHECIMENTO Pág.1 Pág.2	FATURA Pág.1	MANIFESTO Pág.1 Pág.2	ROMANEIO Pág.1	
Nº CERTIFICADO	127418	certeza: 1/1 ocorrências: 1	IMG Pág.1 Pág.2 OCR				---
Nº CRT	001AR224150998	---	---				---
Nº FATURA	0059-00003868	certeza: 1/1 ocorrências: 1	IMG Pág.1 Pág.2 OCR				---
Nº MIC-DTA	16AR231787Y	---	---				ocorrências: 1 Pdf OCR

Figura 13: Certeza e Ocorrências - Método DI -> Docs; Fonte: Elaborado pelo autor.

Outro exemplo, Figura 14, no relatório BatDoc coluna ORIGEM(Certificado de Origem), no campo/linha ENDEREÇO EXP\_1 que corresponde ao endereço do exportador estrangeiro constante na adição 1. A análise do BatDoc foi que se localizou parcialmente tal endereço (4 palavras de um total de 5 = Certeza 4/5) uma vez no documento (ocorrências = 1). Caso não se localize um mínimo definido na configuração de flexibilidade para campos nome e endereço, Figura 8, a análise fica como “Não localizado” (cor vermelha).

CAMPOS	DADOS DI:1519076071	ORIGEM Pág.1 Pág.2	CONHECIMENTO Pág.1 Pág.2	FATURA Pág.1	MANIFESTO Pág.1 Pág.2	ROMANEIO Pág.1
EXPORTADOR_1	RESPONSABILIDAD LIMITADA	certeza 4/6 ocorrências: 2	certeza 5/6 ocorrências: 2	certeza 4/6 ocorrências: 2	certeza 5/6 ocorrências: 1	certeza 4/6 ocorrências: 1
ENDEREÇO Exp_1	DESCARTES 003947 TORTUGUITAS BUENOS AIRES	certeza 4/5 ocorrências: 1				---
NCM_1	39011010	certeza: 1/1 ocorrências: 1				---
INCOTERM_1	CPT	---				---
		---				---

Figura 14: Certeza e Ocorrências - Método DI -> Docs; Fonte: Elaborado pelo autor.



Para cada documento há informações para análise manual do campo da DI, são elas:

<b>IMG</b>	contém link(s) para o(s) recorte(s) do documento onde foi localizada a informação
<b>Pág x PDF</b>	Link para abrir o documento integral
<b>OCR</b>	mostra a(s) linha(s) do arquivo de dados onde o respectivo campo foi localizado.

No exemplo abaixo, Figura 15, para analisar um campo ENDEREÇO EXP\_1, cuja correspondência não foi localizada na Fatura comercial, o Auditor-Fiscal poderia clicar sobre Pág 1 ou Pdf para ter acesso ao documento integral.

CAMPOS	DADOS DI:1612008668	ORIGEM Pág.1 Pág.2	CONHECIMENTO Pág.1 Pág.2	FATURA Pág.1	MANIFESTO Pág.1 Pág.2	ROMANEIO Pág.1
EXPORTADOR_1	[REDACTED]	certeza: 3/3 ocorrências: 1 IMG Pág.1 Pág.2 Pdf OCR	certeza: 3/3 ocorrências: 2 IMG Pág.1 Pág.2 Pdf OCR	certeza: 3/3 ocorrências: 1 IMG Pág.1 Pdf OCR	certeza: 2/3 ocorrências: 1 IMG Pág.1 Pág.2 Pdf OCR	certeza: 3/3 ocorrências: 1 IMG Pág.1 Pdf OCR
ENDEREÇO Exp_1	RUTA 8 KM 60 00 PILAR-BUENOS AIRES ARGENTINA	certeza: 5/8 ocorrências: 1 IMG Pág.1 Pág.2 Pdf OCR	certeza: 6/8 ocorrências: 1 IMG Pág.1 Pág.2 Pdf OCR	Não localizado Pág.1 Pdf OCR	Não localizado IMG Pág.1 Pág.2 Pdf OCR	certeza: 5/8 ocorrências: 2 IMG Pág.1 Pdf OCR
NCM_1	56031130	certeza: 1/1 ocorrências: 1 IMG Pág.1 Pdf OCR				
INCOTERM_1	FCA					

Itens Adição

Comprobante Electrónico  
Factura 0059-00003668  
Original

Ruta # Km 60 - Parque Industrial Pilar  
01620905A Pilar, Buenos Aires  
Tel +54-230 4438100 Fax +54-230 4496321  
IVA Responsable Inscripto

Figura 15: Acessar documento integral - Método DI -> Docs; Fonte: Elaborado pelo autor.

Após verificar a regularidade de informações 'Não localizadas' é possível validar tais campos com um *click*. Assim, esse campo passa para cor verde prosseguindo-se na análise, Figura 16.

Adição_1											
CAMPOS	DADOS DI:1612008668	ORIGEM Pág.1 Pág.2		CONHECIMENTO Pág.1 Pág.2		FATURA Pág.1		MANIFESTO Pág.1 Pág.2		ROMANEIO Pág.1	
EXPORTADOR_1	[REDACTED]	certeza: 3/3 ocorrências: 1	IMG Pág.1 Pág.2 Pdf OCR	certeza: 3/3 ocorrências: 2	IMG Pág.1 Pág.2 Pdf OCR	certeza: 3/3 ocorrências: 1	IMG Pág.1 Pdf OCR	certeza: 2/3 ocorrências: 1	IMG Pág.1 Pág.2 Pdf OCR	certeza: 3/3 ocorrências: 1	IMG Pág.1 Pdf OCR
ENDEREÇO Exp_1	RUTA 8 KM 60 00 PILAR- BUENOS AIRES ARGENTINA	certeza: 5/8 ocorrências: 1	IMG Pág.1 Pág.2 Pdf OCR	certeza: 6/8 ocorrências: 1	IMG Pág.1 Pág.2 Pdf OCR	Não localizado	Pág.1 Pdf OCR	Não localizado	IMG Pág.1 Pág.2 Pdf OCR	certeza: 5/8 ocorrências: 2	IMG Pág.1 Pdf OCR
NCM_1	56031130	certeza: 1/1 ocorrências: 1	IMG X Pág.1 Pdf OCR	++ + = -						---	
INCOTERM_1	FCA	---								---	

Items Adição	
--------------	--

E	Comprobante Electrónico Factura 0059-00003868
Código: 19	Original
<small>Ruta 8 Km 60 - Parque Industrial Pilar 81623985A Pilar, Buenos Aires Tel +54-230 4438100 Fax +54-230 4498321 IVA Responsable Inscripto</small>	

Figura 16: Validar localização negativa - Método DI -> Docs; Fonte: Elaborado pelo autor.

Visando a confiabilidade das informações de localização positiva definiu-se que informações da DI iguais ou menores que 2 caracteres não serão objeto de busca, evitando uma localização falso positiva. Já as informações com mais de 2 caracteres serão objeto de busca com um número máximo de ocorrências esperadas para aquele documento. Quando se atinge esse número máximo de ocorrências, um aviso aparece ao lado da localização positiva para indicar a conferência manual daquela informação da DI.

	Aviso	Indica número de ocorrências de uma informação acima do esperado; Resultante de informações curtas ou cujas especificidades possam comprometer a confiabilidade da localização positiva.
--	-------	---

Com intuito de exemplificar a situação acima configurou-se temporariamente o BatDoc para aplicar busca inclusive sobre informações com 2 caracteres para extração do relatório abaixo. Na Figura 17 verifica-se que o número 20, referente a

quantidade de volumes, foi localizada diversas vezes, dentre as localizações a informação “20 BULTOS” que era a localização esperada, Figura 18. Por isso definiu-se um mínimo de 3 caracteres para informações da DI e uma forma de aviso para número ocorrências acima do esperado.

DI_capa		DADOS DI:1519076071		ORIGEM Pág.1 Pág.2		CONHECIMENTO Pág.1 Pág.2		FATURA Pág.1		MANIFESTO Pág.1 Pág.2		ROMANEIO Pág.1	
VOLUMES (PALLETS)	20	---	---	certeza: 1/1 ocorrências: 5	IMG Pág.1 Pág.2 Pdf OCR	certeza: 1/1 ocorrências: 5	IMG Pág.1 Pág.2 Pdf OCR	certeza: 1/1 ocorrências: 9	IMG Pág.1 Pág.2 Pdf OCR	certeza: 1/1 ocorrências: 3	IMG Pág.1 Pág.2 Pdf OCR	---	---
VMLE US\$	57500	---	---	Não localizado	IMG Pág.1 Pág.2 Pdf OCR	Realize conferência manual, número de ocorrências acima do esperado!	certeza: 1/1 ocorrências: 1	certeza: 1/1 ocorrências: 1	certeza: 1/1 ocorrências: 1	certeza: 1/1 ocorrências: 1	---	---	---
VMCV	60850	---	---	certeza: 1/1 ocorrências: 1	IMG Pág.1 Pág.2 Pdf OCR	certeza: 1/1 ocorrências: 1	certeza: 1/1 ocorrências: 1	---	---	---	---	---	---
FRETE TOTAL	3350.0	---	---	certeza: 1/1 ocorrências: 2	IMG Pág.1 Pág.2 Pdf OCR	certeza: 1/1 ocorrências: 1	certeza: 1/1 ocorrências: 1	certeza: 1/1 ocorrências: 1	certeza: 1/1 ocorrências: 1	---	---	---	---

Figura 17: Aviso de conferência manual - Método DI -> Docs; Fonte: Elaborado pelo autor.

DI_capa		DADOS DI:1519076071		ORIGEM Pág.1 Pág.2		CONHECIMENTO Pág.1 Pág.2		FATURA Pág.1		MANIFESTO Pág.1 Pág.2		ROMANEIO Pág.1	
VOLUMES (PALLETS)	20	---	---	certeza: 1/1 ocorrências: 5	IMG Pág.1 Pág.2 Pdf OCR	certeza: 1/1 ocorrências: 5	certeza: 1/1 ocorrências: 5	certeza: 1/1 ocorrências: 9	certeza: 1/1 ocorrências: 9	certeza: 1/1 ocorrências: 3	certeza: 1/1 ocorrências: 3	---	---
VMLE US\$	57500	---	---	Não localizado	IMG Pág.1 Pág.2 Pdf OCR	Realize conferência manual, número de ocorrências acima do esperado!	certeza: 1/1 ocorrências: 1	certeza: 1/1 ocorrências: 1	certeza: 1/1 ocorrências: 1	certeza: 1/1 ocorrências: 1	---	---	---
VMCV	60850	---	---	certeza: 1/1 ocorrências: 1	IMG Pág.1 Pág.2 Pdf OCR	certeza: 1/1 ocorrências: 1	certeza: 1/1 ocorrências: 1	---	---	---	---	---	---
FRETE TOTAL	3350.0	---	---	certeza: 1/1 ocorrências: 2	IMG Pág.1 Pág.2 Pdf OCR	certeza: 1/1 ocorrências: 1	certeza: 1/1 ocorrências: 1	certeza: 1/1 ocorrências: 1	certeza: 1/1 ocorrências: 1	---	---	---	---

LINHA: 2.010,00 Fecha / (Pág: 1)  
 LINHA: 26-10-2015 .00 USD (Pág: 1)  
 LINHA: 20 BULTOS QUE (Pág: 1)  
 LINHA: -2015 g Notificara (Pág: 1)  
 LINHA: NORTE-ARGENTINA-28-10-2015 RUA FORTE (Pág: 1)

Figura 18: Aviso de conferência manual - Método DI -> Docs; Fonte: Elaborado pelo autor.

### 3.12.2 Navegabilidade em Adição

Em nível de adições da DI, são realizados os batimentos dos campos nome e endereço do exportador/fabricante, NCM e Incoterm utilizando-se os mesmos padrões de buscas vistos na navegabilidade geral. A sistemática pensada para

realizar o batimento em nível de adição foi evitar a redundância de buscas. Logo, se a adição 2 difere apenas a NCM da adição 1 e os demais campos são iguais, será realizado processamento na adição 2 apenas para NCM tendo em vista que para as demais informações já foram realizados os batimentos na adição 1. Caso todos os campos citados de uma adição sejam repetidos, ou seja, já tenham sido objeto de batimento em adições anteriores, mostra-se um sinal de (=).

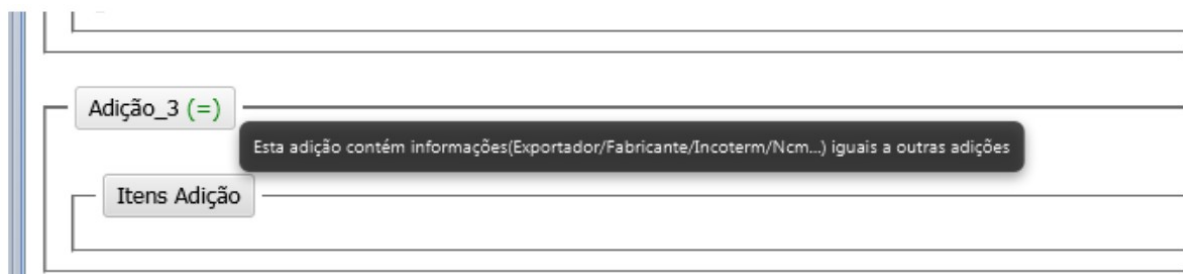


Figura 19: Adição - Método DI -> Docs; Fonte: Elaborado pelo autor.

### 3.12.3 Navegabilidade em Itens da Adição

Os batimentos em nível de itens da adição foram construídos diferentemente dos batimentos em nível de capa da DI e de adição. Isso ocorre tendo em vista que a descrição dos itens na fatura difere muito da DI. Ainda, a descrição muitas vezes omite o código do produto que consta na fatura, utiliza língua ou termos diferentes. Assim, para localização dos itens de cada adição da DI na Fatura comercial construiu-se um algoritmo que analisa 3 variáveis: a quantidade do item, o valor unitário na condição de venda(VUVC) e o total do item em uma mesma linha da fatura, ou seja, num mesmo espaço transversal de linha. Dessa forma, tornou-se possível achar um padrão mínimo para localizar os itens na fatura comercial.

A Figura 20 mostra um exemplo da localização do item 1 da adição 1 na



fatura comercial. O campo análise demonstra que foram localizadas as 3 variáveis utilizadas na busca. A coluna Dados OCR informa a linha do arquivo de dados da fatura onde localizou-se as variáveis.

ITENS na DI						FATURA	
Seq	Unidade	Qtde*	VUCV**	Total do Item(*x**)	Descrição	Análise	Dados OCR
1	peças	100	10,73	1073	FLANGE DE ACOPLAMENTO, DE AÇO LIGA 20MNCR5, DIMENSÕES 100X46MM, TRATAMENTO TÉRMICO DE CEMENTAÇÃO, TÊMPERA E REVENIMENTO UTILIZADA PARA TRANSFERIR A ROTAÇÃO DA TRANSMISSÃO DO MOTOR PARA O EIXO DIFERENCIAL ATRAVÉS DO PINHÃO E COROA DA CAIXA DIFERENCIAL, DE VEÍCULOS AGRÍCOLAS EXCLUSIVO TRATOR ATÉ 100 CVs CÓD 142660	3/3	[0003_142660_ - - BRIDA . PM . 100.000000 . 10.730000 . 1073.00 ](Pag: 1) x

Figura 20: Itens da Adição - Método DI -> Docs; Fonte: Elaborado pelo autor.

Ainda, é possível dar dois *clicks* sobre a linha do OCR para abrir um recorte aproximado da localização da informação na Fatura, conforme demonstra a Figura 21. Nesse caso, pode-se aferir que o item 1 da adição 3 foi localizado e as variáveis(quantidade, VUCV e total do item) conferem com o item 1 da Fatura.

Item	Description	Quantity	Unit Price (EUR)	Amount per Item (EUR)
0001	130195 - ENGRANAJE SAT/LITE EPICIC	200.000000	10.260000	2052.00
0002	130195 - ENGRANAJE SAT/LITE EPICIC	300.000000	10.260000	3078.00

Seq	Unidade	Qtde*	VUCV**	Total do Item(*x**)	Descrição	Análise	Dados OCR
1	peças	200	10,26	2052	ENGRANAGEM PLANETÁRIA EPICICLOIDAL Z 17 DENTES, DE AÇO LIGA, DIMENSÕES M2.75XL32MM, UTILIZADA NOS REDUTORES LATERAIS DO EIXO, FUNÇÃO DE REDUÇÃO DE ROTAÇÃO/TORQUE, DE VEÍCULOS AGRÍCOLAS EXCLUSIVO TRATOR ATÉ 100 CVs CÓD 130195	3/3	[0001_130195_ - - ENGRANAJE . SAT/LITE . EPICIC . 200.000000 . 10.260000 . 2052.00 ](Pag: 1) x
2	peças	300	10,26	3078	ENGRANAGEM PLANETÁRIA EPICICLOIDAL Z 17 DENTES, DE AÇO LIGA, DIMENSÕES M2.75XL32MM, UTILIZADA NOS REDUTORES LATERAIS DO EIXO, FUNÇÃO DE REDUÇÃO DE ROTAÇÃO/TORQUE, DE VEÍCULOS AGRÍCOLAS EXCLUSIVO TRATOR ATÉ 100 CVs CÓD 130195	3/3	[0002_130195_ - - ENGRANAJE . SAT/LITE . EPICIC . 300.000000 . 10.260000 . 3078.00 ](Pag: 1) x
3	peças	9	137,72	1239,48	CONJUNTO COROA E PINHÃO, DE AÇO LIGA, Z12XZ33, DIAMETRO 238,92MM, MONTADOS NA CAIXA DO EIXO DIFERENCIAL, FUNÇÃO DE TRANSMISSÃO DE ROTAÇÃO/TORQUE DE VEÍCULOS AGRÍCOLAS EXCLUSIVO TRATOR ATÉ 100 CVs CÓD 129288	3/3	[0004_129288_ - - PAR . C; NICO . 9.000000 . 137.720000 . 1239.48 ](Pag: 1) x
4	peças	40	127,36	5094,4	CONJUNTO COROA E PINHÃO, DE AÇO LIGA, Z13/Z28, MONTADOS NA CAIXA DO EIXO DIFERENCIAL, FUNÇÃO DE TRANSMISSÃO DE ROTAÇÃO/TORQUE DE VEÍCULOS AGRÍCOLAS EXCLUSIVO TRATOR ATÉ 100 CVs CÓD 144364	3/3	[0005_144364_ - - PAR . C; NICO . 40.000000 . 127.360000 . 5094.40 ](Pag: 1) x

Figura 21: Localização de Itens - Método DI -> Docs; Fonte: Elaborado pelo autor.

Uma decorrência dessa forma de busca é que se a quantidade e o VUCV



forem iguais consequentemente o total do item será igual. Nessa situação provavelmente serão localizadas mais ocorrências para o mesmo item. Por exemplo, os itens 8, 9 e 10 da Figura 22, todos possuem 1020 de quantidade de item, 0,8 de VUCV, logo, total de cada item é o mesmo 816,00 na moeda negociada. Nessa situação, a análise de itens na Fatura ainda é útil pois auxilia como um filtro.

8	peca	1020	0,8	816	BA16443SN - T3 - PUXADOR TD HZD MIST PUXADOR T/D T31 MISTRAL / PUXADOR DE REVESTIMENTO DE PORTA UTILIZADO NA PORTA DE VEICULOS AUTOMOTIVOS CLASSIFICADOS NA POSIÇÃO 8703 NA POSIÇÃO 8703.	3,3	<input type="checkbox"/> 0,8_816.000_000001_1020_UNIDAD_BA16443SN_T3_Puxador_TD_Hzd_Mist_Rto:0005_00030943 [IPag:2] <input checked="" type="checkbox"/> X <input type="checkbox"/> 000002_1020_UNIDAD_BA16440SN_T3_Puxador_DE_Hzd_Mist_T3_Rto:0005.00030943_0,8_816.000_SAP:48949647 [IPag:2] <input checked="" type="checkbox"/> X <input type="checkbox"/> 000003_1020_UNIDAD_BA16441SN_T3_Puxador_DD_Hzd_Mist_T3_Rto:0005.00030943_0,8_816.000_SAP:48949647 [IPag:2] <input checked="" type="checkbox"/> X
9	peca	1020	0,8	816	BA16440SN - T3 - PUXADOR DE HZD MIST PUXADOR D I T31 MISTRAL / PUXADOR DE REVESTIMENTO DE PORTA UTILIZADO NA PORTA DE VEICULOS AUTOMOTIVOS CLASSIFICADOS NA POSIÇÃO 8703.	3,3	<input type="checkbox"/> 0,8_816.000_000001_1020_UNIDAD_BA16443SN_T3_Puxador_TD_Hzd_Mist_Rto:0005_00030943 [IPag:2] <input checked="" type="checkbox"/> X <input type="checkbox"/> 000002_1020_UNIDAD_BA16440SN_T3_Puxador_DE_Hzd_Mist_T3_Rto:0005.00030943_0,8_816.000_SAP:48949647 [IPag:2] <input checked="" type="checkbox"/> X <input type="checkbox"/> 000003_1020_UNIDAD_BA16441SN_T3_Puxador_DD_Hzd_Mist_T3_Rto:0005.00030943_0,8_816.000_SAP:48949647 [IPag:2] <input checked="" type="checkbox"/> X
10	peca	1020	0,8	816	BA16441SN - T3 - PUXADOR DD HZD MIST PUXADOR D D T31 MISTRAL / PUXADOR DE REVESTIMENTO DE PORTA UTILIZADO NA PORTA DE VEICULOS AUTOMOTIVOS CLASSIFICADOS NA POSIÇÃO 8703 NA POSIÇÃO 8703.	3,3	<input type="checkbox"/> 0,8_816.000_000001_1020_UNIDAD_BA16443SN_T3_Puxador_TD_Hzd_Mist_Rto:0005_00030943 [IPag:2] <input checked="" type="checkbox"/> X <input type="checkbox"/> 000002_1020_UNIDAD_BA16440SN_T3_Puxador_DE_Hzd_Mist_T3_Rto:0005.00030943_0,8_816.000_SAP:48949647 [IPag:2] <input checked="" type="checkbox"/> X <input type="checkbox"/> 000003_1020_UNIDAD_BA16441SN_T3_Puxador_DD_Hzd_Mist_T3_Rto:0005.00030943_0,8_816.000_SAP:48949647 [IPag:2] <input checked="" type="checkbox"/> X

Figura 22: Múltiplas localizações de Itens - Método DI->Docs; Fonte: Elaborado pelo autor.


Logo, a resolução nesse caso se faz com a conferência manual a partir do filtro. Clicando-se no botão fechar (X) para itens cuja descrição seja incompatível, concluindo a análise de itens na Fatura o resultado fica conforme a Figura 23.

8	peca	1020	0,8	816	BA16443SN - T3 - PUXADOR TD HZD MIST PUXADOR T/D T31 MISTRAL / PUXADOR DE REVESTIMENTO DE PORTA UTILIZADO NA PORTA DE VEICULOS AUTOMOTIVOS CLASSIFICADOS NA POSIÇÃO 8703 NA POSIÇÃO 8703.	3,3	<input type="checkbox"/> 0,8_816.000_000001_1020_UNIDAD_BA16443SN_T3_Puxador_TD_Hzd_Mist_Rto:0005_00030943 [IPag:2] <input checked="" type="checkbox"/> X
9	peca	1020	0,8	816	BA16440SN - T3 - PUXADOR DE HZD MIST PUXADOR D I T31 MISTRAL / PUXADOR DE REVESTIMENTO DE PORTA UTILIZADO NA PORTA DE VEICULOS AUTOMOTIVOS CLASSIFICADOS NA POSIÇÃO 8703.	3,3	<input type="checkbox"/> 000002_1020_UNIDAD_BA16440SN_T3_Puxador_DE_Hzd_Mist_T3_Rto:0005.00030943_0,8_816.000_SAP:48949647 [IPag:2] <input checked="" type="checkbox"/> X
10	peca	1020	0,8	816	BA16441SN - T3 - PUXADOR DD HZD MIST PUXADOR D D T31 MISTRAL / PUXADOR DE REVESTIMENTO DE PORTA UTILIZADO NA PORTA DE VEICULOS AUTOMOTIVOS CLASSIFICADOS NA POSIÇÃO 8703 NA POSIÇÃO 8703.	3,3	<input type="checkbox"/> 000003_1020_UNIDAD_BA16441SN_T3_Puxador_DD_Hzd_Mist_T3_Rto:0005.00030943_0,8_816.000_SAP:48949647 [IPag:2] <input checked="" type="checkbox"/> X

Figura 23: Itens da Adição - Método DI -> Docs; Fonte: Elaborado pelo autor.

### 3.12.4 Alerta Vermelho

As informações contidas no que se denominou alerta vermelho são produto do método de análise Docs → DI. Conforme abordado em tópico próprio, tal método visa complementar o método de análise DI → Docs. De forma resumida pode-se conceituar da seguinte forma:

	Alerta vermelho	Trata-se de Cnpj, Ncm, Cpf ou nome de empresa que constem nos documentos e não constem na DI; todos os novos padrões incluídos cujo parâmetro foi localizado em alguns dos documentos;
---	-----------------	--

Na Figura 24 o alerta registou que há CNPJ no manifesto rodoviário(MIC/DTA) que não constam na DI.

Método de análise Docs->DI					
Documento	Nome alerta	Parâmetro	OCR	Img	Observações
MANIFESTO_0	CNPJ 1	([^\d]{0-9}[2]{-}\.]{0-9}[3]{-}\.]{0-9}[4]{-}\.]{0-9}[2]{^\d-9}([^\d]{0-9}[2]{-}\.]{0-9}[3]{-}\.]{0-9}[4]{-}\.]{0-9})	nseguro010091370017220000-1722	IMG	---
MANIFESTO_0	CNPJ 2	([^\d]{0-9}[2]{-}\.]{0-9}[3]{-}\.]{0-9}[4]{-}\.]{0-9}[2]{^\d-9}([^\d]{0-9}[2]{-}\.]{0-9}[3]{-}\.]{0-9}[4]{-}\.]{0-9})	-67000164	IMG	---
MANIFESTO_0	CNPJ 3	([^\d]{0-9}[2]{-}\.]{0-9}[3]{-}\.]{0-9}[4]{-}\.]{0-9}[2]{^\d-9}([^\d]{0-9}[2]{-}\.]{0-9}[3]{-}\.]{0-9}[4]{-}\.]{0-9})	-67000164	IMG	---

Figura 24: Alerta Vermelho; Fonte: Elaborado pelo autor.

Logo, tal alerta funcionou plenamente pois foi capaz de identificar que havia um CNPJ o qual não foi objeto de declaração na DI mas constava em um documento instrutivo. Nessa situação, o Auditor-Fiscal realiza a devida apuração da inconsistência. Tais alertas gerados são muito relevantes pois podem auxiliar na identificação de nomes de empresas e CNPJ de terceiras empresas que constem nos documentos podendo ser reais adquirentes não declarados. Passando o mouse

sobre IMG, Figura 25, exibe-se o recorte do documento onde foi localizada tal informação, especificamente tal CNPJ é do transportador e este não é declarado em campo próprio na DI.

Método de análise Docs->DI					
Documento	Nome alerta	Parâmetro	OCR	Img	Observações
MANIFESTO_0	CNPJ 1	([^\d]{0-9}[2][^\d]{0-9}[3][^\d]{0-9}[3][^\d]{0-9}[4][^\d]{0-9}[2][^\d]{0-9})	nseguro01009137001722	IMG	---
MANIFESTO_0	CNPJ 2	([^\d]{0-9}[2][^\d]{0-9}[3][^\d]{0-9}[3][^\d]{0-9}[4][^\d]{0-9}[2][^\d]{0-9})	567000164	IMG	---
MANIFESTO_0	CNPJ 3	([^\d]{0-9}[2][^\d]{0-9}[3][^\d]{0-9}[3][^\d]{0-9}[4][^\d]{0-9}[2][^\d]{0-9}[3][^\d]{0-9}[3][^\d]{0-9}[4][^\d]{0-9})		IMG	---

MANIFESTO_0 CNPJ 3	
10 Rol de Contribuyentes	11 Placa de Placa
567000164	
12 Marca y número	13 Capas
14 Año	15 Placas
2010	
23 Nº carta de porte	24 Aduana de destino
25 Moneda	26 Origen de las mercancías

Figura 25: Parâmetros Alerta Vermelho; Fonte: Elaborado pelo autor.

Com intuito de testarmos o funcionamento dos novos padrões de localização do método análise Docs → DI incluímos um novo padrão, Figura 26, cujas informações foram nome: NCM monitorada, parâmetro: 56031130 e uma observação.

Identificação(nome)	Parâmetro	Observações
NCM monitorada	56031130	Irregularidade DI anterior xx - Identificar g/m <sup>2</sup> e material constitutivo na verificação física, documentos e descrição DI

Obs.: A Identificação e Parâmetro são obrigatórios para salvar o registro.  
 Obs.: Ao Salvar os dados incluídos, novos campos em branco são criados para novas inserções.

OK

Figura 26: Tela Incluir Novos Padrões(Parâmetros); Fonte: Elaborado pelo autor.

Ao aplicar o BatDoc gerou-se um alerta vermelho com o parâmetro



monitorado conforme Figura 27. Já mencionou-se em tópico próprio que tal ferramenta, além de buscar qualquer conjunto alfanumérico nos documentos instrutivos, é capaz de servir como uma base de conhecimento pessoal para o trabalho no curso do despacho. Assim, seria possível incluir uma observação para uma NCM ou uma empresa com sanção administrativa, dentre outras possibilidades.

Método de análise Docs->DI					
Documento	Nome alerta	Parâmetro	OCR	Img	Observações
MANIFESTO_0	CNPJ 1	{([^\d]{0-9}){2}[-.]{0-9}{3}[-.]{0-9}{3}[-\w]{0-9}{4}[-\w]{0-9}{2}[-\d]{0-9}){([^\d]{0-9}){2}[-.]{0-9}{3}[-.]{0-9}{3}[-\w]{0-9}{4}[-\w]{0-9})	nseguro0100913700 [REDACTED] 7001722	IMG	---
MANIFESTO_0	CNPJ 2	{([^\d]{0-9}){2}[-.]{0-9}{3}[-.]{0-9}{3}[-\w]{0-9}{4}[-\w]{0-9}{2}[-\d]{0-9}){([^\d]{0-9}){2}[-.]{0-9}{3}[-.]{0-9}{3}[-\w]{0-9}{4}[-\w]{0-9})	[REDACTED] 567000164	IMG	---
MANIFESTO_0	CNPJ 3	{([^\d]{0-9}){2}[-.]{0-9}{3}[-.]{0-9}{3}[-\w]{0-9}{4}[-\w]{0-9}{2}[-\d]{0-9}){([^\d]{0-9}){2}[-.]{0-9}{3}[-.]{0-9}{3}[-\w]{0-9}{4}[-\w]{0-9})	[REDACTED] 57000164	IMG	---
ORIGEM_0	NCM monitorada	56031130	5603.11.30 Las demas	IMG	Irregularidade DI anterior xx - Identificar g/m <sup>2</sup> e material constitutivo na verificação física, documentos e descrição DI

Figura 27: Parâmetros Alerta Vermelho; Fonte: Elaborado pelo autor.

Tal localização ocorreu na primeira folha do certificado de origem (ORIGEM\_0) para o parâmetro de teste incluído (NCM monitorada) cujo valor foi a NCM 56031130. Pode-se utilizar o link IMG para aferição do ponto aproximado do documento onde tal informação foi localizada, Figura 28.



Método de análise Docs->DI					
Documento	Nome alerta	Parâmetro	OCR	Img	Observações
MANIFESTO_0	CNPJ 1	[(^0-9)[0-9]{2}[^\\.]?[0-9]{3}[^\\.]?[0-9]{3}[^\\.]?[0-9]{4}[^\\.]?[0-9]{2}[^0-9])((^[0-9]{0-9}{2}[^\\.]?[0-9]{3}[^\\.]?[0-9]{3}[^\\.]?[0-9]{4}[^\\.]?[0-9]{2}[^0-9])	nseguro01009137001 [REDACTED] 37001722	IMG	---
MANIFESTO_0	CNPJ 2	[(^0-9)[0-9]{2}[^\\.]?[0-9]{3}[^\\.]?[0-9]{3}[^\\.]?[0-9]{4}[^\\.]?[0-9]{2}[^0-9])((^[0-9]{0-9}{2}[^\\.]?[0-9]{3}[^\\.]?[0-9]{3}[^\\.]?[0-9]{4}[^\\.]?[0-9]{2}[^0-9])	[REDACTED] 67000164	IMG	---
MANIFESTO_0	CNPJ 3	[(^0-9)[0-9]{2}[^\\.]?[0-9]{3}[^\\.]?[0-9]{3}[^\\.]?[0-9]{4}[^\\.]?[0-9]{2}[^0-9])((^[0-9]{0-9}{2}[^\\.]?[0-9]{3}[^\\.]?[0-9]{3}[^\\.]?[0-9]{4}[^\\.]?[0-9]{2}[^0-9])	[REDACTED] 67000164	IMG	---
ORIGEM_0	NCM monitorada	56031130		IMG	Irregularidade DI anterior xx - Identificar g/m² e material constitutivo na verificação física, documentos e descrição DI

ORIGEM_0 NCM monitorada		
5. Medio de transporte previsto VIA TERRESTRE		
Nº DE ORDEN	5. CODIGOS NCM	10. DENOMINACION
1	5603.11.30	LAS DEMAS / POLIPROPILENO TECIDO NAO TRAZIDO EN PROCESO SS
		PESO BRUTO: [REDACTED]
		TOTAL: [REDACTED]
		MARCAS: [REDACTED]

Figura 28: Parâmetros Alerta Vermelho; Fonte: Elaborado pelo autor.

### 3.12.5 Links Web Automáticos

O relatório BatDoc possui alguns links web na parte superior visando agilidade na análise do despacho. É possível a inclusão/alteração/exclusão de links web personalizando o relatório BatDoc, Figura 29.

Nome do Link	Endereço Web
google	http://www.google.com.br
Cuit/AFIP-Argentina	https://seti.afip.gob.ar/padron-puc-constancia-internet/Con...
RFB	http://idg.receita.fazenda.gov.br/
Antt	https://appweb1.antt.gov.br/scff/conPlaca.asp
Vicomex	https://www.siscomex.gov.br/vicomex/public/index.jsf

Obs.: Trata-se de uma forma de personalização dos links web disponíveis no relatório BatDoc.  
 Os links aqui incluídos ficam dispostos na opção (+)links do relatório  
 Ex.: Nome do Link: RFB - Endereço Web: http://idg.receita.fazenda.gov.br

OK

Figura 29: Tela Incluir Novos Links; Fonte: Elaborado pelo autor.

Os primeiros 10 links web incluídos ficam dispostos numa barra/linha superior ao relatório denominado “Consultas Úteis”, Figura 30.



Figura 30: Links relatório BatDoc; Fonte: Elaborado pelo autor.

A partir do 11º link os links estão dispostos no botão “+”, inclusive o acesso ao manual do BatDoc.

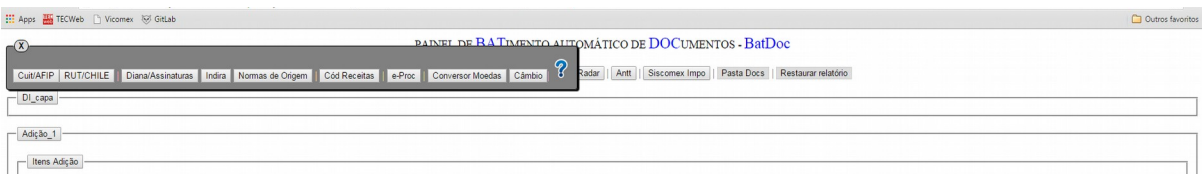


Figura 31: Links (+) relatório BatDoc; Fonte: Elaborado pelo autor.

Seguindo a linha da agilidade pretendida, elaboramos 5 links automáticos. Com eles, por meio de um click, o usuário tem acesso a determinada página, conforme segue:

- Google Exp/Google Imp: abre página com pesquisa no Google pelo nome do importador ou exportador estrangeiro, Figura 32;

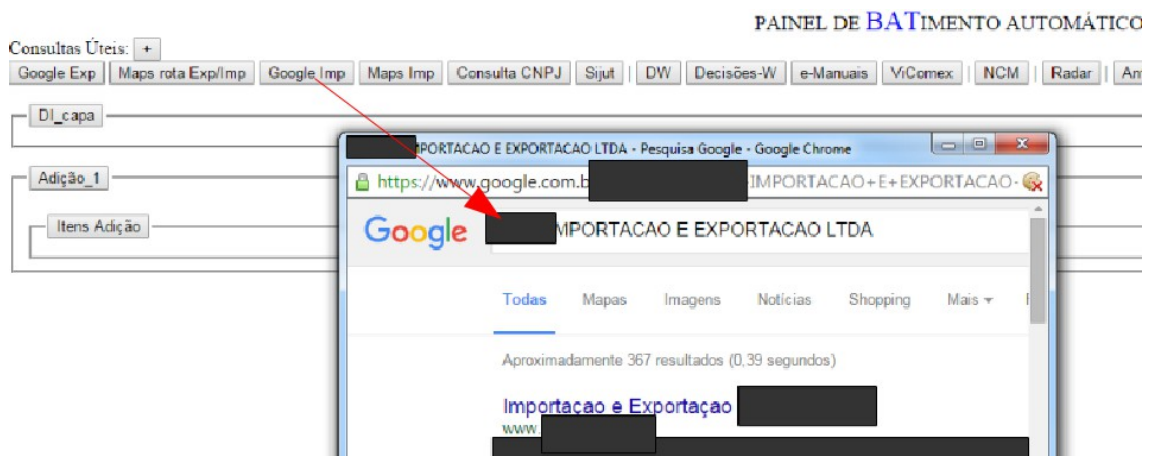


Figura 32: Link automático pesquisa Google; Fonte: Elaborado pelo autor.

- Antt: abre a página de consulta de habilitação do veículo rodoviário de transporte internacional para a placa do veículo rodoviário informada na DI, Figura 33;

CAMPOS	DADOS DI: 1519076071	ORIGEM Pág. 1 Pág. 2	CONHECIMENTO Pág. 1 Pág. 2	FATURA Pág. 1																						
PLACA	4697			---																						
PESO BRUTO Kg	25500	<table border="1"> <thead> <tr> <th colspan="2">Dados do Veículo</th> </tr> </thead> <tbody> <tr> <td>Placa:</td> <td>4697</td> </tr> <tr> <td>Tipo:</td> <td>IMI-REBOQUE</td> </tr> <tr> <td>Modelo:</td> <td>FECHADA</td> </tr> <tr> <td>Chassi/Motor:</td> <td>DF1513YY5150922</td> </tr> <tr> <td>CMT:</td> <td>3 eixeladas</td> </tr> <tr> <td>CCU:</td> <td>3 toneladas</td> </tr> <tr> <td>Marca:</td> <td>RANDON</td> </tr> <tr> <td>Eixos:</td> <td>3</td> </tr> <tr> <td>Ano:</td> <td>2000</td> </tr> <tr> <td>TARA:</td> <td>9 toneladas</td> </tr> </tbody> </table>		Dados do Veículo		Placa:	4697	Tipo:	IMI-REBOQUE	Modelo:	FECHADA	Chassi/Motor:	DF1513YY5150922	CMT:	3 eixeladas	CCU:	3 toneladas	Marca:	RANDON	Eixos:	3	Ano:	2000	TARA:	9 toneladas	/1 IMG /1 IMG /1 IMG
Dados do Veículo																										
Placa:	4697																									
Tipo:	IMI-REBOQUE																									
Modelo:	FECHADA																									
Chassi/Motor:	DF1513YY5150922																									
CMT:	3 eixeladas																									
CCU:	3 toneladas																									
Marca:	RANDON																									
Eixos:	3																									
Ano:	2000																									
TARA:	9 toneladas																									
PESO LÍQUIDO Kg	25000			/1 IMG /1 IMG /1 IMG																						
VOLUMES (PALLETS)	20			/1 IMG /1 IMG /1 IMG																						

Figura 33: Link automático pesquisa Antt; Fonte: Elaborado pelo autor.

- Maps Imp: abre página Google Maps no ponto do endereço do importador, possibilitando de imediato os modos de visualização satélite/streetView, Figura 34.

Figura 34: Link automático Google Maps StreetView; Fonte: Elaborado pelo autor.

- Maps rota Exp/Imp: abre página Google Maps traçando a rota do endereço do exportador estrangeiro até o endereço do importador, Figura 35;

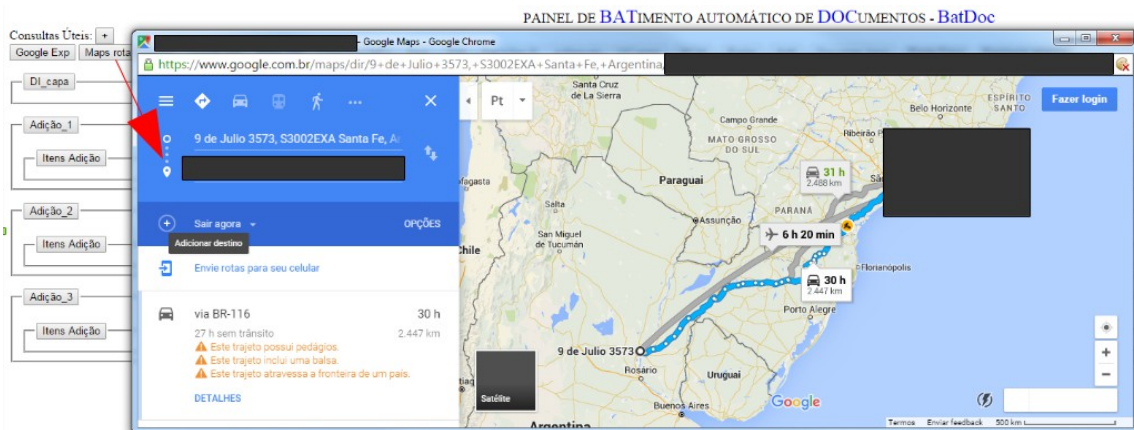


Figura 35: Link automático Google Maps rota; Fonte: Elaborado pelo autor.

## 4 TRABALHOS DERIVADOS

### 4.1 BatDoc

#### 4.1.1 Aplicabilidade no Gerenciamento de Risco

Atualmente o art. 19, §2º, da IN SRF nº 680/2006, estabelece uma faculdade para a Coana dispensar a vinculação do dossiê com os documentos instrutivos à DI parametrizada em canal verde. A Coana por meio da PORTARIA COANA Nº 30, DE 02/2015 dispensou a vinculação de tal dossiê.

No futuro, decidindo-se pela anexação digital dos documentos instrutivos em DI parametrizada em canal verde, possibilita-se um alargamento da análise de risco realizada sobre as operações de importação. Se por um lado o importador precisaria realizar a anexação para todos os despachos de importação, por outro, a recepção seria automática, o importador tem o dever de guarda dos documentos e, ainda,



haveria possibilidades aprimoramento da seleção. Por exemplo, dirimir dúvidas de conformidade documental em tempo de seleção influenciando a decisão de redirecionamento ou não de uma DI para conferência de canal vermelho.

Sem dúvidas, realizar uma análise documental no canal verde durante o tempo disponível e a quantidade de DI's nos lotes de parametrização é visivelmente inviável. Nesse cenário, os métodos de análise de dados não estruturados apresentados, ou mais especificamente o BatDoc, podem fornecer um subsídio relevante. O fato de já estar sendo desenvolvido dentro do Aniita facilita muito esse desafio.

Por outro lado, há desafios a serem enfrentados para esse tipo de aplicação tendo em vista o processamento massivo de dados. Por exemplo, o tempo de processamento do *OCR-Abbyy*. Tal tempo consome mais de 80% do tempo de processamento do BatDoc.

Em relação ao método de análise DI → Docs imaginemos um lote de parametrização, onde podem ser eleitos os campos da DI e documentos instrutivos mais importantes para processamento do BatDoc para fins de seleção. A partir da análise das inconsistências seria possível tomar a decisão pelo redirecionamento.

Já no método de análise Docs → DI, os padrões de CNPJ, NCM, nome de empresas e eventuais novos padrões incluídos gerariam o alerta vermelho de cada DI possibilitando uma verificação manual das informações omitidas na DI.



#### **4.1.2 Aplicabilidade em Revisões Aduaneiras de DI**

A aplicabilidade dos métodos de análise de dados objeto deste trabalho em revisões aduaneiras parecem ser promissoras do ponto de vista teórico. Para isso é necessário o uso massivo de dados tendo em vista a necessidade de *download* de inúmeros dossiês para fins de seleção/fiscalização.

Na perspectiva do BatDoc não utiliza-se uso massivo de dados visto que a aplicação inicial no piloto dar-se-á no âmbito do curso do despacho de importação. Assim, o processamento é realizado apenas para a DI objeto de conferência aduaneira nos canais vermelho e amarelo, além de retificações de DI, sob responsabilidade do Auditor-Fiscal.

Assim como já citado na Aplicabilidade no Gerenciamento de Risco, os principais desafios para aplicação em revisões aduaneiras também são a anexação obrigatória dos canais verdes e o processamento do *OCR*. O ideal seria o próprio ViComex após a anexação do dossiê digital efetivar o processamento dos documentos por meio de um *OCR*, isso seria similar ao e-processos onde o sistema disponibiliza uma versão pesquisável do PDF.

No método de análise DI → Docs é possível fazer a verificação de qualquer informação de um conjunto de DIs para com os documentos que a instruem. Tal abordagem tende a revisões aduaneiras por alguma desconformidade da DI para com seus documentos, por exemplo, análise da NCM no certificado de origem do Mercosul.

Já na abordagem do método de análise Docs → DI seria possível fazer a verificação de qualquer padrão ou parâmetros específicos em documentos



digitalizados que sejam relevantes para a seleção/fiscalização. Por exemplo, utilizar tal método para reconhecer CPF, CNPJ, NCM, nome de empresa, ou qualquer informação específica em qualquer dos documento instrutivo de um conjunto de Dis.

#### **4.2 Potencial dos Métodos de Análises**

Genericamente considerado o método de análise dado estruturado → dado não estruturado possibilita sua aplicação sempre que se fizer necessário aferir uma declaração eletrônica perante documentos digitalizados.

Já o método de análise dado não estruturado → dado estruturado é de possível aplicação sempre que se quiser efetivar buscas de padrões(parâmetros) em documentos digitalizados. Logo, qualquer informação alfanumérica pode ser incluída para fins de buscas em documentos digitalizados.

### **5 TRABALHOS FUTUROS**

Pretende-se continuar desenvolvendo e aperfeiçoando os métodos de análise de dados não estruturados com aplicação no curso do despacho. Para tanto, utilizar-se principalmente dos *feedbacks* da fase piloto agregando as melhorias necessárias e possíveis.

### **6 CONCLUSÃO**

Os desafios enfrentados durante a análise estritamente documental podem ser ao menos mitigados com a ferramenta apresentada. Tal solução não substitui o





trabalho do Auditor-Fiscal, mas sim, traz um subsídio que visa minimizar o tempo da análise meramente documental direcionado-se a atuação fiscal para análises complexas relativas ao despacho. Os *feedbacks* oriundos do período de piloto serão extremamente relevantes para melhorar as possibilidades de auxílio dessa solução.

Logo, 30 itens de mercadoria podem ser localizados automaticamente na fatura, NCM's são localizadas no Certificado de Origem, uma maior certeza de que os documentos não possuem CNPJ ou nomes de empresas estranhos a negociação comercial, enfim uma solução para gerar agilidade na análise estritamente documental.

O presente trabalho apresenta os métodos de análise de dados não estruturados e possíveis aplicações no contexto aduaneiro visando, principalmente, instigar mais projetos nesse sentido.

Não se trata de um método infalível visto limitações quanto à forma e à qualidade da anexação, às diferentes formas de prestação de informação na DI e nos documentos e à devida extração do texto pelo *OCR*. Em que pese as diversas variáveis que influenciam no funcionamento do BatDoc, tem-se resultados favoráveis para maior parte das análises de DI na unidade do autor.

Os avanços tecnológicos são inerentes ao comércio mundial. Por exemplo, inicia-se com Brasil e Argentina o projeto do certificado de origem digital (COD), quando o COD estiver em pleno funcionamento seus dados estarão dispostos em dados estruturados não se aplicando necessariamente os métodos de análise de dados não estruturados. Sob tal perspectiva os dados do COD poderiam ser analisados e dispostos em mais uma aba do projeto Aniita, sendo mais uma fonte de



informações estruturadas. Ainda nesse contexto o BatDoc tem aplicação para os demais documentos(Fatura, MIC, TIFF, conhecimento de embarque e romaneio) apresentados para análise no despacho.

Se há uma frase que pode resumir o sentimento em relação ao trabalho apresentado é a seguinte: *“Não há nada que seja maior evidência de insanidade do que fazer a mesma coisa dia após dia e esperar resultados diferentes”*, atribuída a Albert Einstein.

#### **REFERÊNCIAS:**

VICOMEX - Manual de utilização do Visão Integrada e Módulo Anexação de Documentos – 2015. Disponível em <[http://portal.siscomex.gov.br/informativos/biblioteca-de-arquivos/Manual\\_VICOMEX.pdf](http://portal.siscomex.gov.br/informativos/biblioteca-de-arquivos/Manual_VICOMEX.pdf)>. Acesso em 10 de setembro de 2016.

COUTINHO, G. L.. Aniita – uma abordagem pragmática para o gerenciamento de risco aduaneiro baseada em software. Prêmio de Criatividade e Inovação da RFB, 2012.

JAMBEIRO FILHO, J. E. S. – Sisam - Inteligência Artificial no Sistema de Seleção Aduaneira por Aprendizado de Máquina, 2015

TESSAROLO, P. H. - Os dados estruturado e não estruturados. Disponível em <[http://web.unipar.br/~seinpar/2015/\\_include/artigos/Pedro\\_Henrique\\_Tessarolo.pdf](http://web.unipar.br/~seinpar/2015/_include/artigos/Pedro_Henrique_Tessarolo.pdf)>. Acesso em 03 de setembro de 2016.

TESSERACT – Disponível em <<https://github.com/tesseract-ocr/tesseract>>, 2006. Acesso em 04 de setembro de 2016.

ABBYY. Abby Hot Folder. Disponível em <<http://help.abbyy.com/FineReader/FineReader12/English/HotFolder/HotFolder.htm>>. Acesso em 27 de agosto de 2016.

REGEX – Regular Expression. Disponível em <<http://docs.oracle.com/javase/tutorial/essential/regex/>>. Acesso em 28 de agosto de 2016.

RFB – Site da Secretaria da Receita Federal do Brasil. Disponível em <<http://idg.receita.fazenda.gov.br/>>. Acesso 26 de agosto de 2016.

## ANEXO 1

### Mapeamento de campos versus documentos para os quais são realizados batimentos automatizados de informações

Campos	CO Mercosul	Conhecimento			Fatura	Manifesto		Romaneio
		CRT	BL-HBL	AWB		MIC/DTA	TIF/DTA	
<b>CAPA DI</b>								
Nº CERTIFICADO	X							
Nº CRT/BL-HBL-AWB		X	X	X		X	X	
Nº Fatura	X	X			X			
Nº MIC/DTA – TIF/DTA						X	X	
Nº ROMANEIO								X
IMPORTADOR	X	X	X	X	X	X	X	X
ENDEREÇO Imp	X	X	X	X	X	X	X	X
CNPJ	X	X	X	X	X	X	X	X
ADQUIRENTE	X	X	X	X	X	X	X	
ENDEREÇO Adq	X	X	X	X	X	X	X	
CNPJ Adq	X	X	X	X	X	X	X	
TRANSPORTADOR		X	X	X		X	X	
EMBARCAÇÃO			X					
CIDADE EMBARQUE	X	X	X	X				
DATA EMBARQUE		X	X	X				
QTD VOLUMES		X	X	X	X	X	X	X
CIDADE ENTREGA		X	X	X				
PESO BRUTO		X	X	X	X	X	X	X
PESO LÍQUIDO		X			X	X	X	X
VMLE		X			X	X	X	
VMCV		X			X			
FRETE TOTAL		X	X	X	X	X	X	
FRETE EXTERNO		X	X	X	X			
FRETE NACIONAL		X						
SEGURO TOTAL		X	X	X	X	X	X	
PLACAS VEÍCULO						X		
<b>ADIÇÃO</b>								
EXPORTADOR_	X	X	X	X	X	X	X	X
ENDEREÇO Exp_	X	X	X	X	X	X	X	X
FABRICANTE_					X			
ENDEREÇO Fab_					X			
INCOTERM_		X	X	X	X			
NCM_	X						X	
<b>ITENS DA ADIÇÃO</b>								
QUANTIDADE *					X			
VUCV **					X			
SUBTOTAL (* x **)					X			