



SEARCH-IN-A-BOX

APPLIANCE DE INTELIGÊNCIA

SOLUÇÃO INTEGRADA



Search-in-a-box Appliance de Inteligência Solução Integrada

Inteligência, combate à lavagem de dinheiro, combate à fraude financeira, combate ao crime cibernético ou qualquer outra forma de aplicação da Lei, demandam recursos e métodos de altíssima tecnologia.

Embora os agentes e investigadores são a peça chave no ambiente investigativo, as quantidades e complexidades das diversas fontes de informação disponíveis no mundo moderno demandam ferramentas tecnológicas que acelerem e apoiem o processo cognitivo na busca de pistas e respostas.

Computadores apreendidos, flash drives, telefones celulares, redes sociais, bases de dados ou conjunto de dados governamentais e judiciais de movimentações bancárias e de movimentações financeiras anômalas, interceptações telefônicas e bases de dados contendo informações criminais são apenas alguns poucos exemplos do que um analista precisa enfrentar no dia-a-dia das investigações criminais.



Um grande desafio e sua solução.

Em 26 anos de experiência no mercado brasileiro e 2 anos de experiência no mercado americano, a Via Appia foi capaz de captar e endereçar alguns importantes desafios do setor de inteligência:

- Alta complexidade dos softwares e equipamentos;
- Dificuldade para coletar e buscar nas diversas fontes de informação;
- Baixa integração entre as soluções próprias e de terceiros;
- Infraestrutura de hardware insuficiente ou mal dimensionada.

Assim, a Via Appia imbuíu-se na tarefa de criar uma solução que ao mesmo tempo fosse o mais simples possível de usar e trouxesse a maior completude e possibilidade de agregação de recursos. O usuário tem a opção de escolher escalabilidades funcionais e/ou de capacidade a qualquer momento.

Para suprir essa necessidade, a Via Appia criou o Appliance de Inteligência Search-in-a-box, ou de forma mais simples, SIAB, como vem sendo chamado pelos usuários. Trata-se de uma solução integrada de software e hardware com um módulo central baseado em conectores de conteúdo diversos, indexação e busca inteligente em contextos através do uso de ontologias. Enquanto os usuários podem se ater apenas na coleta de informações estruturadas, não estruturadas e semiestruturadas em fontes híbridas e nas funções de busca avançada, eles também podem acrescentar, de acordo com sua demanda, funções de análise de vínculos, transcrição de áudio para texto, resolução de identidade e mais uma série de outros módulos de software.

Além de embarcar e integrar várias ferramentas próprias com outras líderes de Mercado como IBM, Tableau, Basis, dentre outros, o SIAB permite a integração com bases de dados e ferramentas próprias das agências ou entidades conveniadas. A integração pode ir do simples acesso a bancos de dados até a integração através de serviços.

Para vencer os desafios descritos acima, o SIAB foi desenvolvido seguindo premissas importantes de facilidade, poder de processamento, completude de funções, alta integração e capacidade de consumir dados híbridos com transformação ou preservação dos formatos originais, conforme demanda do usuário. Características básicas do SIAB:

- Plug'n play;
- Vários módulos (ferramentas ou soluções de software) que funcionam como "peças de lego", onde o próprio usuário monta sua configuração;
- O resultado gerado por um módulo é insumo para o processamento do outro;
- Os dados podem ser compartilhados indefinidamente;
- O SIAB já traz a integração básica entre as ferramentas;
- Outras formas de interação podem ser configuradas ou desenvolvidas pelos usuários ou sob demanda.

*Search-in-a-box,
simplicidade e
poder.*

SEARCH-IN-A-BOX



SOFT OU HARDWARE APPLIANCE

A solução completa proposta pela Via Appria envolve o conceito mais estrito de appliance, ou seja, uma solução integrada de hardware e software, no entanto, o SIAB pode ser fornecido em forma de soft appliance, ou seja, uma máquina virtual do appliance original que pode ser instalada em servidores do próprio usuário.

É chamada solução completa pelo fato de ser plug'n play e trazer a configuração otimizada para os módulos embarcados suplanta certas dificuldades técnicas que deverão ser vencidas na instalação em outro servidor.

SEGURANÇA DAS INFORMAÇÕES

O SIAB possui uma estrutura de segurança da informação extremamente robusta, principalmente no que tange à infecção(malware) e invasão. Ao entrar no mercado americano, o SIAB foi submetido a uma avaliação pela acreditada empresa Coalfire Systems e teve aprovação imediata, sem a necessidade de ajustes ou desenvolvimento extra. Desde 2012, quando foi criado, até o presente momento não foi reportada qualquer falha na segurança.

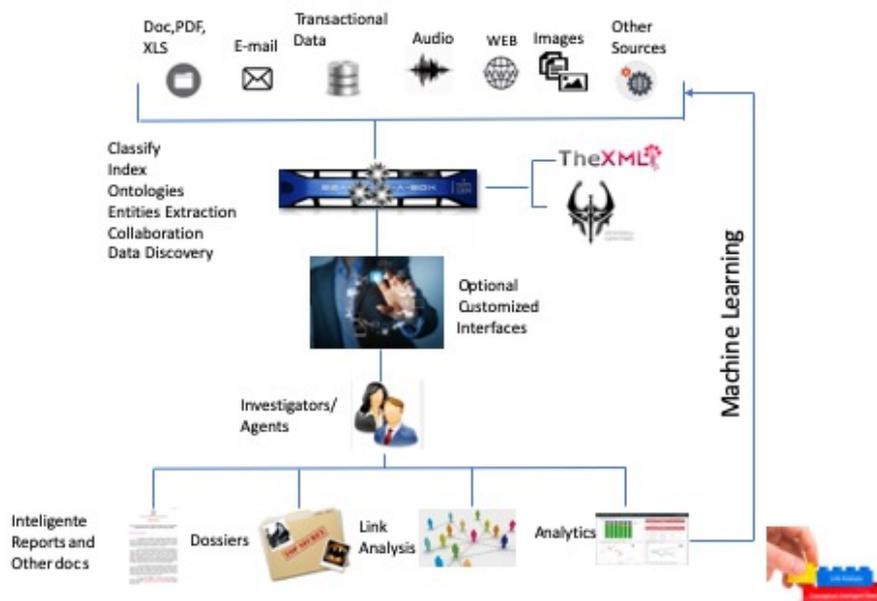
CAPTURA E INDEXAÇÃO DE INFORMAÇÕES

Como vários módulos de software com diferentes funções estão disponíveis, considerando ainda as soluções e informações do próprio usuário, a solução oferece um conjunto muito grande de crawlers ou

Integração sem perder a flexibilidade de uma instalação distribuída, o usuário escolhe entre um appliance de software (VM) ou hardware.

conectores de conteúdo. Esses crawlers, ou conectores de conteúdo, para facilitar, são capazes de capturar virtualmente qualquer tipo de dado aos quais os usuários terão acesso. Para suprir alguma deficiência, o SIAB traz um framework de desenvolvimento de conectores. Exemplos de conectores disponíveis abaixo:

- Hadoop e outros bancos NoSQL
- Email (exchange, Notes, etc.)
- Sistemas de arquivo (docx, xlsx, pdf, etc.)
- Internet and Intranet
- Bancos de Dados Relacionais (JDBC)
- Redes Sociais
- Mídia Rica (transformação do áudio e do vídeo em texto pesquisável e OCR)
- Reconhecimento Facial
- Captura de outras informações utilizando Web Services e APIs



Funcionalidade e capacidade sob demanda, escale como quiser.

BIG DATA, DATA LAKE, DATA AT REST

Desde seu lançamento no mercado, um dos diferenciais do SIAB sempre foi a capacidade de tratamento de Big Data, ou seja, de vastas quantidades de informação. Para tal, no que tange à pesquisa e recuperação de informações com o uso ou não de ontologias, garante-se um resultado sempre em sub segundos, independente do tamanho da fonte dos dados.

O SIAB sustenta, de acordo com a configuração dos módulos de software, os três níveis de repositório, Big Data , Data Lake e Data at Rest. Este último é considerado para as informações que estão em evidência na investigação em um dado momento, melhor dizendo, uma fotografia do estado das informações em um lapso de tempo. Essa coleção de dados é tornada estática, pelo menos temporariamente, para evitar que os usuários ou sistemas automatizados possam alterá-las, como seria o caso de uma mensagem do Twitter que pode ser alterada a qualquer momento, invalidando uma linha da investigação.

Os dados em movimento, ou streaming, são o exemplo claro de Big Data e são monitorados enquanto vão acontecendo, ou seja, mensagens de uma rede social ou transações bancárias podem gerar alertas pré-programados para os agentes e analistas.

Enquanto as duas formas de repositório ou captura acima acontecem, um Data Lake permanece ao dispor do usuário para complementar ou embasar a investigação. O Data Lake normalmente é formado por dados de referência, de outras investigações ou outras fontes de uso geral. Também pode ser alimentado com resultados notáveis da captura dos dados em movimento e dos dados investigados enquanto ainda estavam no status de Data at Rest.

O Data Lake gerenciado pelo SIAB segue à risca as características envolvidas nesse conceito, dentre as quais podemos citar a captura e manutenção de dados híbridos, ou seja, estruturados, não estruturados e semiestruturados e em seus formatos nativos. Assim, temos no Data Lake arquivos do word, excel, pdf, dados de bancos de dados relacionais, vídeos, áudios e outros. Os módulos de software que compõem as “peças de lego” da solução acessam e compartilham tais informações, de acordo com sua necessidade.

BUSCA INTELIGENTE EM CONTEXTOS: O CORAÇÃO DO SEARCH-IN-A-BOX

A ideia da criação do SIAB, conforme já dito acima, se baseou nas dificuldades enfrentadas pelos usuários na utilização de ferramentas avançadas para o processo investigativo. As duas primeiras ferramentas foram a busca inteligente em contextos e a análise de vínculos. Por busca inteligente em contextos, entende-se uma

A informação que não pode ser encontrada é como um baú de tesouro no fundo do mar, tem grande valor, mas não tem qualquer utilidade.

ferramenta de busca capaz de entender conceitos e sua ocorrência dentro de contextos.

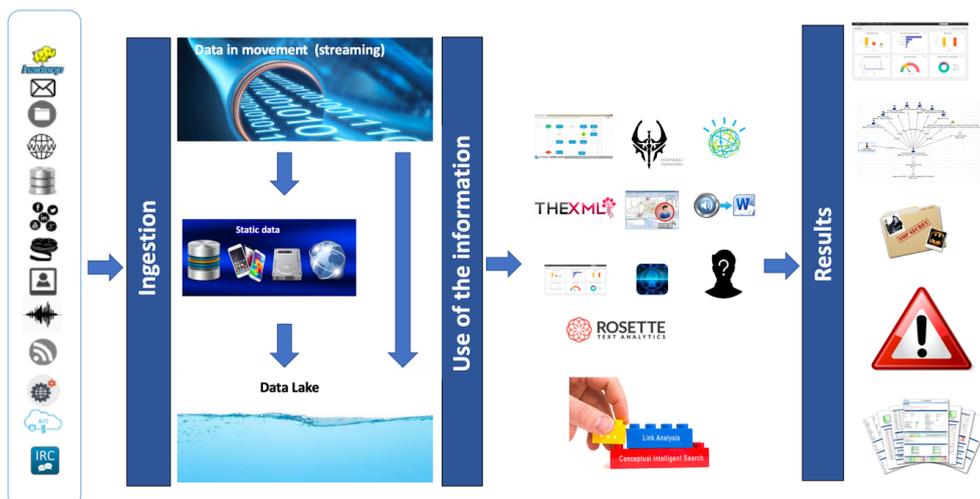
Para definição e manutenção de contextos o SIAB usa o software VA TheXML e para proceder à busca contextual, utiliza opcionalmente o VA Heimdall ou o IBM Watson Explorer Deep Analytics. Essas duas ferramentas interagem com o TheXML e, a partir daí, conseguem entender, por exemplo, que cocaína é uma droga, que possui sinônimos como pó e farinha e que tem termos específicos, como por exemplo, as substâncias utilizadas para o refino da droga, como éter etílico e acetona. A partir desse entendimento, o SIAB consegue descobrir, como exemplo, duas pessoas falando sobre éter etílico, sendo que no contexto existe uma grande possibilidade de se estar falando do refino da cocaína.

De tão lógico, é redundante dizer que quando as pessoas buscam algo, o mais importante é encontrar. Quem pesquisa na Internet talvez já tenha se acostumado, mas o que temos nos motores de busca da internet hoje é uma total falta de funcionalidade, seja lá qual for o buscador. Faz-se uma pesquisa por "cocaína", como foi o exemplo anterior e o que teremos serão milhões de resultados planos, contendo apenas a palavra ou um pouco mais que isso.

O uso das funcionalidades de busca em uma investigação, similar a qualquer outra aplicação de motores de busca, visa encontrar o maior número de informações importantes no menor resultado possível, ou seja, com o mínimo de ruídos na busca. São os chamados índices de precisão e recall. Deixar informações importantes no "fundo do lago" ou simplesmente perdê-la em uma lista de resultados sem fim é o que vai determinar conseguir ou não encontrar um ator ou um comportamento criminoso.

Garantir o melhor resultado na busca e descoberta de informações é a parte mais significativa da proposta do Search-in-a-box. As integrações e tratamento das informações gravitam em torno da busca por ontologias e dos outros recursos não menos importantes embarcados. Todos os módulos de software contam com a possibilidade de utilizar as informações já processadas pelos outros módulos e assim por diante.

*Big Data e Data
Lakes –
Repositórios
Híbridos.*



RECURSOS DE IA COM USO DE APRENDIZADO DE MÁQUINA

O SIAB possui para os módulos de busca inteligente em contextos e mineração de textos recursos de aprendizado de máquina que permitem ao usuário criar anotadores e definir padrões que são submetidos a um processo de aprendizagem, permitindo a descoberta automática de informações correlatas. Além dessa possibilidade, todos os resultados gerados pelos módulos do SIAB podem ser retroalimentados para aumentar a inteligência do Sistema.

ONTOLOGIAS E VOCABULÁRIOS CONTROLADOS

O módulo VA TheXML é uma solução completa para criação e manutenção de vocabulários controlados e dicionários científicos, englobando Listas, Anéis de Sinônimos, Taxonomias, Thesaurus e Ontologias. Foi totalmente desenvolvido seguindo a norma ISO 24.965-1 e 24.965-2. Dessa forma, provê os relacionamentos descritos na norma e, ainda de acordo com ela, permite a criação de outros relacionamentos entre a terminologia.

O TheXML foi criado em 1994 e vem seguindo a evolução tanto das normas quanto de outros padrões internacionais. O uso dos dicionários/vocabulários técnicos não está limitado ao módulo de busca inteligente em contextos, mas também aos mais diversos módulos do SIAB, incluindo aprendizado de máquina, ou machine learning.

Análise de Vínculos, Busca Inteligente em Contextos, Analíticas de Texto, Speech-to-text (fala para texto), BI, Reconhecimento Facial, Resolução de Identidade, Análise Preditiva, Aprendizado de Máquina e IA.

ANÁLISE DE VÍNCULOS

Para análise de vínculos, o SIAB utiliza dois módulos que são escolhidos de acordo com o nível de aplicação a ser dada. Os dois módulos são o IBM i2 e o IBM i2 EIA – Enterprise Insight Analysis. Enquanto que no i2 podemos selecionar diversas entidades como pessoas, veículos, empresas, telefones, endereços, dentre outros e gerar gráficos que mostrem o relacionamento entre elas, no EIA podemos trabalhar diversos tipos de dados transformando-os em inteligência abrangente e acionável.

Trabalhando de forma integrada, ou seja, quando embarcado no appliance, o módulo de análise de vínculos pode receber as entidades extraídas das fontes de dados coletadas e indexadas pelo módulo de busca e plotar os diagramas de análise. O módulo IBM EIA oferece um gama muito grande de possibilidades, incluindo o próprio i2 Analyst's Notebook e iBase.

RESOLUÇÃO DE IDENTIDADE

O módulo de resolução de identidade (IBM Recommendation Engine) visa responder duas perguntas básicas: quem é quem? e quem conhece quem? Para a primeira pergunta, a solução varre todos os dados estruturados disponíveis e através de recursos de aprendizado de máquina, procura descobrir as mais diversas formas com as quais estão identificados o mesmo indivíduo.

Para descobrir quem conhece quem, a solução, utilizando-se dos mesmos dados, procura a coincidência do uso de informações e recursos, por exemplo, pessoas que utilizaram os mesmos números de telefone, que disseram residir no mesmo lugar, que utilizaram os mesmos cartões de milhagens e muito mais.

As informações descobertas podem ser utilizadas nos outros módulos para gerar resultados em forma de gráficos ou tabulares. O módulo de busca inteligente em contextos, por sua vez, faz a extração de informações estruturadas dos documentos e informações não estruturadas, dando ao usuário a possibilidade de expandir a análise e conseguir maior precisão.

INTELIGÊNCIA DE NEGÓCIOS E ANÁLISE PREDITIVA

O SIAB transforma dados não estruturados em metadados e clusters que podem ser utilizados em BI e outros módulos analíticos melhorando a revocação e precisão do sistema.

Quando se fala de grandes ou enormes fontes de informação muitos aspectos da mesma acaba sendo difícil, quando não impossível, de ser analisada e muita coisa fica por ser descoberta. A busca inteligente, a análise de vínculos e a resolução de entidades respondem às mais importantes perguntas e auxiliam a descobrir os indivíduos e suas ações, mas grandes quantidades de informação podem guardar em si padrões e modelos que devem ser enfrentados de outras formas.

Separamos aqui dois módulos distintos, mas que estão intimamente interligados. O primeiro, baseado na ferramenta Tableau, permite a navegação e análise em grandes quantidades de dados estruturados gerando dashboards e resultados analíticos. É importante lembrar aqui que os dados não estruturados podem, até certo ponto, serem estruturados pelo módulo principal de busca inteligente em contextos.

Se todos os módulos anteriores auxiliaram na coleta, busca, descoberta e análise das informações, outro módulo muito importante é o de Análise Preditiva. O módulo de análise preditiva é responsável por analisar todas as informações coletadas e produzidas e “prever” acontecimentos futuros. Ou seja, delinear as possibilidades de repetição ou não de padrões no comportamento de atores e grupos criminosos ou ainda de os mesmos mudarem seus comportamentos se movendo para outras regiões ou mudando seu modus operandi.

MÍDIA RICA

Dentre as opções que o SIAB permite embarcar, estão módulos de tratamento de mídia rica. Esses módulos proveem a capacidade de transcrever áudio em texto (speech-to-text), extrair e transcrever o áudio dos arquivos de vídeo, OCR de imagens e de PDF de imagens, permitindo inclusive a transcrição dos PDFs de imagem para texto. Em todos esses casos as informações extraídas podem ser alvo da extração de entidades, descoberta de dados, transformação de partes não estruturadas em metadados e muitas outras funcionalidades fornecidas pelos demais módulos do SIAB.

Ao transcrever o áudio e o áudio de vídeos para texto, o SIAB indexa e permite a busca inteligente nos conteúdos. Dentre os áudios transcritos, podemos exemplificar com as interceptações telefônicas,

As funcionalidades de mídia rica do SIAB incluem speech-to-text, de arquivos de áudios ou áudios de vídeos, OCR de imagens e PDFs de imagem, Audio fingerprint e mais.

áudios de call center, vídeos de inquéritos policiais, dentre vários outros exemplos.

Ao efetuar uma busca, o SIAB posiciona o usuário nos locais que contém os termos pesquisados, marcando o tempo no qual eles acontecem. O usuário então clica no termo e o SIAB toca o áudio do que está sendo dito a partir daquele momento. O tempo marcado serve para que o usuário possa fazer referência aos momentos importantes do áudio em seu relatório.

Existe um grande número de idiomas possíveis, sendo que português ou inglês são embarcados automaticamente dependendo do país que faz a aquisição. Também estão disponíveis russo, mandarim, árabe, e outros sob consulta.

RECONHECIMENTO FACIAL

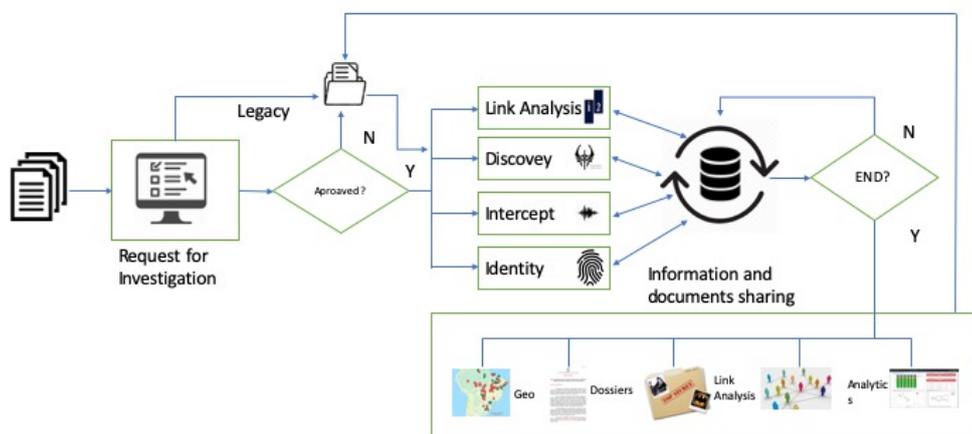
O módulo de reconhecimento facial que está disponível sob demanda é o software FaceVACS da empresa Americana Cognitec. O FaceVACS permite a criação de uma base de dados de fotografias de pessoas e, ao receber a submissão de uma foto a ser pesquisada, ordena as várias possibilidades de combinações de acordo com o percentual de acerto provável.

GERENCIAMENTO AVANÇADO DE CASOS

O SIAB utiliza o IBM BAW para automatizar e orquestrar o processo investigativo. Esse módulo reúne funcionalidades de ECM e BPM para criar o ambiente de gestão de todos os processos administrativos e investigativos. Os processos são desenhados pelo próprio usuário ou pelos consultores da Via Appia e todos os envolvidos nas investigações têm o resultado de seu trabalho nos diversos módulos armazenados no repositório de documentos ou dados do Sistema.

Dessa forma, pessoas que possuem a permissão de acesso suficiente podem ter acesso ao resultado do trabalho dos colegas e proceder às complementações e atualizações. Com o uso do Gerenciamento Avançado de Casos, todas as informações utilizadas no decorrer dos processos investigativos ficam organizadas em um único local e garantem um trabalho colaborativo entre os agentes/investigadores.

O usuários podem integrar soluções de terceiros para adquirir mais e mais poder. Sem limites na tecnologia.



OUTROS MÓDULOS – MAIS PODER

Mesmo que não embarcados, sistemas e soluções do próprio cliente podem ser integrados aos módulos do SIAB. Isso permite criar um ambiente operacional totalmente integrado, trazendo para os agentes/analistas tudo o que precisam sem a necessidade de enfrentar um ambiente repleto de pequenas ilhas de dados a serem integradas manualmente, ou até mesmo trabalharem isoladamente.

Bases de dados judiciais, governamentais ou de agências de aplicação da lei podem ser integradas através dos conectores de conteúdos ou através das opções fornecidas pelo SIAB. Todas as fontes de informação ou soluções podem ser integradas com uso das seguintes técnicas:

- Compartilhamento de informações armazenadas em containers;
- Web Services, SOAP ou REST;
- IBM Integration BUS;
- SIAB's APIs;
- Integração desenvolvida sob demanda.

VIA APPIA INFORMATICA EIRELI (BR)

SIA Trecho 1 – Lotes 630-870 – Bl. 3
Salas 01-05 – Ed. Praça Capital
71200-012 – Brasília – DF – Brasil
www.viaappia.com.br
comercial@viaappia.com.br

VIA APPIA CORP (US)

2000 Technology Drive, Suite 450
15219 – Pittsburgh – PA - EUA
+1 412 573-2099
www.viaappia.com
comercial@viaappia.com