

Qualidade de Dados no Modelo de Transformação de Informações como Suporte à Projeções Situacionais

Valdir Amancio Pereira Junior, *Programa de Pós-Graduação em Ciência da Informação, Universidade Estadual Paulista "Júlio de Mesquita Filho"*, Marília, São Paulo, Brazil, valdir.pereira@unesp.br

Abstract—Situational Awareness (SAW) refers to the level of consciousness that an individual or team holds about a situation. In the criminal domain, SAW failures can induce human operators to make mistakes in decision making and to generate imminent risks to life and heritage. In addition, critical domains with dynamic situations are prone to information quality problems, especially when the source of the data is human beings. Considering the nature of the information and the context that is involved, the information may be incomplete, outdated, inconsistent or influenced by cultural and stress factors. Thus, the objective of this study is to discuss the role of data quality as a relevant part in a model of information transformation and how it can relate to inference and representation processes. Partial results indicate the real need for adherence and continuous evaluation of data quality to assist SAW and situational projections. It is also hoped that the model can support the processing of information and quality analysis, resulting in information with lower dimensionality, higher quality indexes and more representative to help the human being.

Keywords—Data Quality, Situational Awareness, Situational Projections

Resumo—Consciência Situacional (SAW) refere-se ao nível de consciência que um indivíduo ou equipe detém sobre uma situação. No domínio criminal, falhas da SAW podem induzir operadores humanos a cometer erros na tomada de decisão e gerar riscos iminentes à vida e patrimônio. Além disso, domínios críticos com situações dinâmicas são propensos a problemas de qualidade da informação, especialmente quando a fonte dos dados são seres humanos. Considerando a natureza das informações e o contexto que estão envolvidos, as informações podem ser incompletas, ultrapassadas, inconsistentes ou influenciadas por fatores culturais e de estresse. Desta maneira, o objetivo deste trabalho é discutir o papel da qualidade de dados como parte relevante em um modelo de transformação de informação e como esta pode se relacionar com processos de inferência e representação. Resultados parciais indicam a real necessidade da aderência e avaliação contínua da qualidade dos dados para auxiliar a SAW e projeções situacionais. Espera-se que também, que o modelo possa dar suporte ao processamento de informações e análise de qualidade, resultando em informações com menor dimensionalidade, índices de melhor qualidade e mais representativos para auxiliar o ser humano.

Palavras-chave—Qualidade de Dados, Consciência Situacional, Projeções Situacionais

I. INTRODUÇÃO

Consciência Situacional (Situational Awareness – SAW) é um processo cognitivo importante para tomadores de decisão em diversas áreas e se refere ao nível de consciência que um indivíduo ou equipe detém sobre uma situação. Especificamente, SAW diz respeito à percepção da presença e

disposição de entidades de interesse em um ambiente, digital ou analógico, à compreensão do significado e importância de suas ações individuais e coletivas no espaço-tempo e à projeção de seus status em um futuro próximo [1].

No domínio criminal, SAW é um fator crucial para, com base em dados coletados de múltiplas fontes, processados e devidamente representados, revelar tendências, ocorrências e a possibilidade de riscos iminentes. Uma SAW limitada pode comprometer a compreensão sobre o estado real de um ambiente de interesse, levando a uma má tomada de decisão, podendo resultar em consequências para vidas, patrimônios e até o meio ambiente. Embora o acesso a informações sobre as situações não possa garantir SAW e a melhor tomada de decisão de um humano, uma informação de qualidade pode colaborar com tomadores de decisão a fim de eliminar a incerteza, manter um conhecimento superior e atingir eventualmente o nível de projeção situacional [2].

A projeção situacional é a capacidade de gerar uma estimativa sobre um conjunto de possibilidades futuras de uma situação, baseado nos elementos e padrões atuais de uma situação no presente. A capacidade de projetar situações é essencial pois SAW não é apenas sobre o que está acontecendo no momento presente, mas também sobre eventos e ações que vão acontecer a seguir. Considerando que o humano não consegue influenciar no presente de uma situação, apenas no seu futuro, tornando assim os elementos de uma situação atual insumos para a projeção situacional e compreensão de estados futuros [3].

Adquirir e manter uma SAW é ainda mais desafiador ao considerar o contexto de dados criminais no Brasil, levando em conta o grande volume e heterogeneidade dos mesmos. Outro fator crítico é a qualidade dos dados e suas respectivas fontes, em sua maioria de inteligência humana, neste caso, relatos de crimes às centrais de atendimento das forças de segurança. Tipicamente tais dados são incompletos, desatualizados, inconsistentes e influenciados por fatores culturais, afetando os processos informacionais e computacionais que coletam, organizam, recuperam, processam e representam os dados e estimulam a SAW do analista [4], [5].

Considerando o contexto e desafio apresentados, a Ciência da Informação (CI) se posiciona como fator essencial para viabilizar e organizar a comunicação entre diversas ciências e disciplinas, de acordo com seu caráter interdisciplinar, Saracevic [6] já aponta que CI pode melhorar as formas como as informações são utilizadas. Suplementado por Borko [7], que

traz a CI como forma de investigar as formas, comportamentos e fluxos informacionais, além de considerar o processamento da informação, visando otimizar a acessibilidade e usabilidade. Questões essenciais ao considerar a qualidade da informação, uma vez que este conceito é subjetivo e diretamente ligado ao domínio estudo, sendo que a avaliação dos índices de qualidade da informação irá depender de parâmetros e padrões das ciências associadas ao domínio, não apenas a informação em si.

Sobre essa visão interdisciplinar e embasados na CI, diversos trabalhos vêm buscando formas de superar as problemáticas de coletar, armazenar, transformar, processar, inferir, representar, recuperar e apresentar dados relacionados à domínios de emergência [4], [5], [8], [9], [10].

Sendo que tais processos afetam diretamente os dados que são consumidos por um humano que irá tomar decisões, agregando diversas ciências junto a CI, como a Ciência da Computação, Psicologia, Ciência Cognitiva e disciplinas como Fusão de Dados e Informações, Ontologias, Qualidade da Informação, Recuperação de Informações entre outras.

Desta maneira, este trabalho traz uma discussão sobre como a qualidade de dados e informações pode colaborar com a estruturação de um modelo de transformação e representação de informações em suporte à obtenção de projeções situacionais. Para tal, é discutida a influência da qualidade no arranjo tecnológico, envolvendo também a fusão de informações, SAW, ontologias e interface de usuário. Para isso, o trabalho é estruturado em seções na seguinte sequência: (2) Consciência Situacional, (3) Qualidade de Dados, (4) A Qualidade na Estruturação do Modelo de Transformação e (5) Considerações Finais.

II. CONSCIÊNCIA SITUACIONAL

Endsley [11] traz uma das definições mais amplamente aceitas hoje para SAW, colocando-a como a percepção das entidades que compõem um ambiente, considerando as variáveis de tempo e espaço, buscando a compreensão do significado e contexto em que se encontra, assim como realizar a projeção dentro de um futuro próximo [4], [11].

Endsley [11] também posiciona a SAW como uma rotina, um ciclo, um processamento de informações dirigida pela aquisição e uso destas, estando assim, presente em diversas situações, desde atividades operacionais e técnicas, que exigem ou altíssimo nível de concentração e estado de SAW, até atividades comuns ao dia-a-dia.

Neste trabalho será adotado o modelo de Endsley [11], pois este modelo é dirigido por dados e informações, que sofrem ações de sistemas externos, possibilitando uma correlação sistêmica [4]. Este modelo é composto por três níveis, com elementos que se relacionam, considerando, ações externas de outras tarefas e sistemas.

O nível um representa a percepção de entidades, estados e características, junto aos fatores que delimitam o ambiente. O nível dois está relacionado a associação entre os elementos percebidos no nível um, caracterizando a compreensão da situação que é observada, isto envolve a interpretação de um grande volume de dados e informações. Enquanto o nível três é

o nível mais alto de SAW, visando alcançar a projeção de uma situação, ou seja, um estado futuro das informações, baseado no nível dois.

III. QUALIDADE DE DADOS

O conceito de qualidade é variável e muitas vezes subjetivo, alinhado às necessidades, ações e objetivos de cada domínio, caracterizando-se como parâmetros para que o produto ou resultado esteja de acordo com as necessidades. Como o produto analisado está livre de problemas e/ou falhas, ele permite que os fatores dependentes do produto analisado sejam executados com êxito. Como a qualidade é inerente a vários domínios e pode ser aplicada a dados, informações e outros produtos, os problemas e as dimensões da qualidade também são mutáveis [12].

Abordar diretamente a qualidade dos dados visa alinhá-lo com as necessidades do domínio de que eles vêm ou são aplicados. Os dados que não descrevem fielmente os elementos do ambiente real interferem na efetividade dos Sistemas de Informação (SI) e podem gerar falhas no próprio sistema. Em atividades do mundo real, a qualidade dos dados é um elemento que possibilita a SAW e a tomada de decisões em um domínio [4], [8].

Para [13], novas questões e problemas relacionados à quantidade e qualidade dos dados estão surgindo em um ambiente de Tecnologias de Informação e Comunicação (TIC) em rápida mutação. Os autores também fazem uma comparação entre as tarefas que o sistema de informação realizou e seus requisitos, destacando o uso de tarefas comuns no passado, como problemas de classificação e processamento de dados homogêneos, e como os requisitos e as tarefas evoluíram, se tornando cada vez mais complexas, mesmo lidando com a SAW.

Considerando essas mudanças nos sistemas de informação, juntamente com as mudanças no meio ambiente, o espectro informativo se expande consideravelmente. Existem diferentes fontes de dados de diferentes tipos de sensores, tais como sensores físicos, robôs e algoritmos de mineração, sensores de imagem e sensores humanos (ou seja, de inteligência humana). Com a crescente quantidade de dados heterogêneos sendo capturados e armazenados nas mais diversas coleções de dados, lidar com todos esses dados em toda a sua amplitude não é mais o único problema a ser resolvido. Considerar, avaliar e melhorar a qualidade também se tornaram uma necessidade e preocupação.

De acordo com English [14], nas últimas décadas foi destacado um novo conceito, a interferência da qualidade dos dados na execução e na capacidade dos SI. A má qualidade impacta negativamente várias facetas do ambiente de informação, a partir do custo do SI, através da representação através da interface gráfica e afetando níveis mais críticos, como a construção da SAW e tomada de decisão.

IV. QUALIDADE DE DADOS COMO SUPORTE AO MODELO DE TRANSFORMAÇÃO

Considerando que um modelo representa um arranjo organizado de técnicas, tecnologias e até mesmo outros modelos, com intuito de atender as necessidades informacionais

de um domínio, aplicação ou problemática. Ao considerar a estruturação de um modelo que prevê a transformação e representação de informações que irão auxiliar o humano na construção de projeções situacionais, este vai de encontro com alguns pontos da CI apontados por Borko (1968) [7], como a coleta, organização, armazenamento, recuperação, interpretação e transformação da informação.

Esta seção se dedica a apresentar o início de um modelo que prevê auxiliar a transformação e representação de projeções informacionais, que possam auxiliar o humano no processo de construção de SAW e tomada de decisão. Para isso, são necessários alguns módulos capazes de lidar com as informações da devida maneira, neste caso, capazes de lidar com dados provenientes de humanos, problemas de qualidade intrínsecos aos dados, informações semânticas e SAW.

A Figura 1, apresenta o modelo, destacando em vermelho a avaliação de qualidade, junto aos demais componentes de Coleta de Dados, Transformação, Fusão, Representação das Situações, Inferências de Projeções Situacionais e o Modelo de Representação.

Alguns trabalhos já discutem e trazem modelos e técnicas capazes de lidar com os pontos relacionados. Explorando o uso da fusão semântica de informações, ontologias capazes de representar índices de qualidade, interfaces que possibilitam a devida representação das informações e o uso de modelos de consciência situacional.

Alguns destes também começam a considerar processos que possibilitem a projeção situacional ou a organização de demais processos para que possam auxiliar o humano a atingir a projeção situacional, como apresentado por Pereira Junior, Pereira e Botega [8]. Porém tais trabalhos consideram a avaliação dos índices de qualidade como um processo interno, que cria insumos para os processos de inferência, representação e recuperação, não dando a devida base à qualidade dos dados, sendo que essa deve estar presente em todos os processos que manipulam a informação.

Botega et al. [4] também apresenta um modelo de fusão de informações (Quantify), capaz de fazer análises sintática e semântica de informações, junto a avaliação de qualidade visando melhorar SAW. Sendo que a qualidade aparece como uma das bases do modelo, porém não é discutido qual a sua relevância para o modelo e como esta influenciou no arranjo tecnológico que foi pensado, além de não considerar a capacidade de produzir, representar e avaliar informações sobre um futuro, as projeções informacionais, dando indícios de que é possível sanar problemas informacionais ao considerar a qualidade de dados dentro dos processos que manipulam informações.

Com isso o presente trabalho adota técnicas, processos e modelos já discutidos como a aplicação da Human-driven Brazilian Crime Ontology (URBANITY), ontologia capaz de representar situações criminais que tem como principal fonte relatos humanos sobre crimes, junto a índices de qualidade e a metodologia Information Quality Assessment Methodology in the Context of Emergency Situational Awareness (IQESA), desenvolvida com o foco em avaliar e representar os índices de qualidade para que sejam aplicados junto ao processo de avaliação de informações. Agregando também técnicas

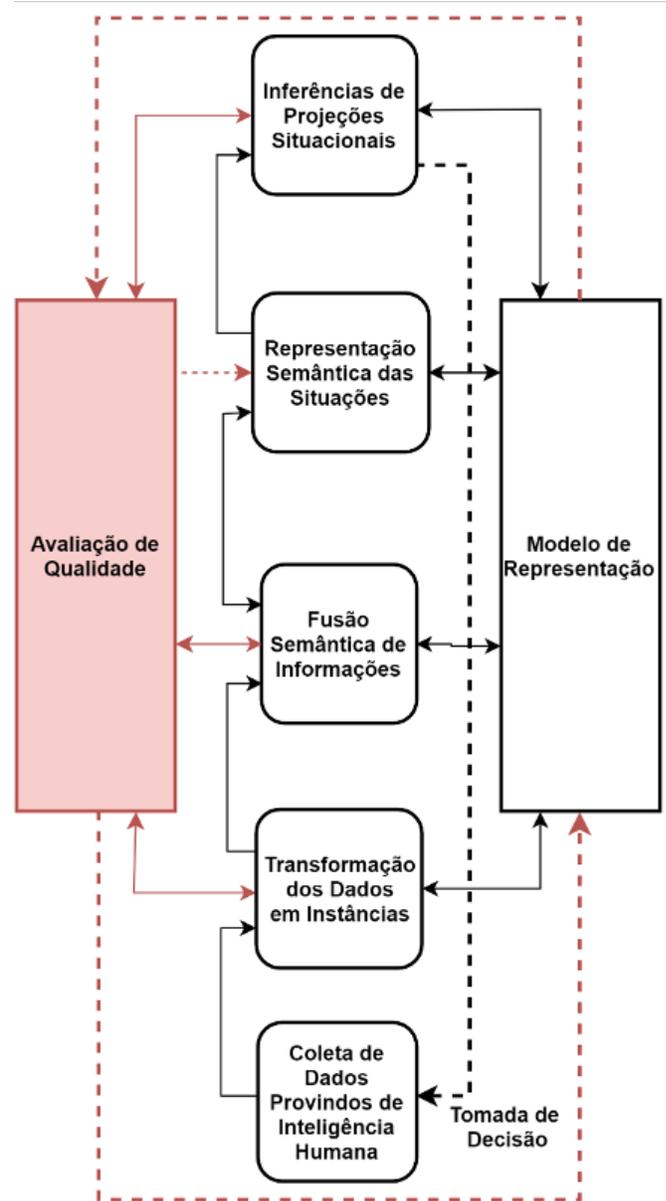


Fig. 1. Modelo de transformação e representação de projeção situacionais. Fonte: Elaborado pelo autor.

de fusão semântica de informações que consideram critérios semânticos e de qualidade, aplicando este durante os processos de busca e associação de informações [4], [5], [8].

Considerando os apontamentos feitos por Melo, Botega e Santarém Segundo [5], que traz os problemas de qualidade não relacionando apenas aos dados diretamente, mas considera também a estrutura provida para sua publicação, sendo que este pode gerar falhas e dificuldades no acesso, assim como impossibilitar o devido consumo, resultados que embasam e direcionam as discussões deste trabalho para trazer a qualidade como processo e ativo e participativo dentro de todo o modelo. Junto com Pereira Junior, Pereira e Botega [8], onde afirmam que a melhoria na qualidade dos dados pode influenciar positivamente a SAW, permitindo assim que humanos atinjam a

projeção situacional através do consumo dos dados de melhor qualidade.

Este modelo se apropria de recomendações de modelos anteriores e supera limitações dos mesmos, onde é possível notar que a qualidade não está apenas relacionada a avaliação e representação de índices de qualidade, ela permeia todo o ciclo de vida das informações dentro de um modelo, servindo como parâmetros de busca, capacidade de inferência dentro de uma ontologia e até mesmo estimulam o usuário final em alguns domínios. Por isso a necessidade de considerar a mesma a partir do momento de estruturação de um modelo que siga uma abordagem informação perante as bases da CI.

Com a intuito de validar a discussão e alcançar os primeiros passos na direção de concretizar um modelo de transformação e representação de informações totalmente calçado por índices de qualidade, foi elaborado o modelo demonstrado na Figura 1, associando todos os possíveis módulos do modelo e destacando como os índices de qualidade e o módulo de qualidade, estarão diretamente associados a todos. Neste caso considerando desde a coleta dos dados, o processamento e transformação realizados por um processo de fusão de informações semânticas, a avaliação de qualidade e um modelo de representação de dados, que no caso será aplicado a URBANITY, devido ao domínio considerado.

Ao analisar o modelo proposto, é possível notar a presença contínua, desde a transformação dos dados coletados em instância seguindo o Modelo de Representação, aqui representado por Transformação de Dados em Instâncias, os quais os dados já são avaliados, assim, o processo de fusão já se torna capaz de inferir e produzir novas informações baseado no conteúdo, semântica e índices de qualidade das informações. Os resultados produzidos por esse módulo são novamente avaliados, podendo trazer novas dimensões e métricas ainda não consideradas. As novas informações e seus respectivos índices de qualidade, são representados aqui na “URBANITY” que novamente tem a capacidade de interagir com a avaliação de qualidade para além de poder representar semanticamente os índices de qualidade, através de vocabulários específicos, pode reavaliar alguns índices de qualidade, como a consistência e completude, agora sobre uma visão semântica, possibilitando inferir propriedades e dimensões ainda não consideradas, enriquecendo mais o resultado da fusão.

V. CONSIDERAÇÕES FINAIS

Uma vez que a qualidade da informação permeia todo o modelo, sendo considerada desde o momento de sua estruturação, torna possível criar ciclos informacionais totalmente cientes de qualidade e gerar um acompanhamento mais detalhe, ou seja, se torna possível identificar onde ocorrem falhas de processamento, inferência, transporte, recuperação e apresentação das informações que sendo aplicadas no modelo.

Outro fator resultante da aplicação de qualidade dentro do modelo é apresentação de informações mais confiáveis ao humano e logo, uma SAW mais bem construída, levando assim o humano ao nível de projeção situacional de uma maneira mais ágil.

REFERÊNCIAS

- [1] M. M. Kokar and M. R. Endsley, “Situation awareness and cognitive modeling,” *IEEE Intelligent Systems*, vol. 27, no. 3, pp. 91–96, 2012.
- [2] G. L. Rogova and E. Bosse, “Information quality in information fusion,” in *2010 13th International Conference on Information Fusion*. IEEE, 2010, pp. 1–8.
- [3] J. Roy, “From data fusion to situation analysis,” in *Proceedings of Fourth International Conference on Information Fusion*, vol. 2, 2001.
- [4] L. C. Botega, A. C. M. de Oliveira, V. A. P. Junior, J. F. Saran, L. Z. Ladeira, G. M. C. Pereira, and S. Isotani, “Quantify: An information fusion model based on syntactic and semantic analysis and quality assessments to enhance situation awareness,” in *Information Quality in Information Fusion and Decision Making*. Springer, 2019, pp. 563–586.
- [5] J. O. d. S. F. MELO, L. C. BOTEGA, and J. E. S. SEGUNDO, “Metodologia de avaliação de qualidade para dados conectados:: Brapci 2.0,” *Informação & Tecnologia; v. 4, n. 2 (2017): Informação & Tecnologia-Especial Enancib 2017: 80-101*, vol. 24, no. 2, pp. 101–80, 2018.
- [6] T. Saracevic, “Interdisciplinary nature of information science,” *Ciência da informação*, vol. 24, no. 1, pp. 36–41, 1995.
- [7] H. Borko, “Information science: what is it?” *American documentation*, vol. 19, no. 1, pp. 3–5, 1968.
- [8] V. A. P. Junior, G. M. C. Pereira, and L. C. Botega, “Towards a process for criminal semantic information fusion to obtain situational projections,” in *The Human Position in an Artificial World: Creativity, Ethics and AI in Knowledge Organization*. Ergon-Verlag, 2019, pp. 51–72.
- [9] G. M. C. PEREIRA, J. N. SILVA, V. A. PEREIRA JUNIOR, Á. C. M. d. OLIVEIRA, and L. C. BOTEGA, “Processo de análise quantitativa de eventos criminais utilizando abordagem semântica:: Brapci 2.0,” *Informação & Tecnologia; v. 4, n. 1 (2017): Informação & Tecnologia-Especial I WIDaT; 69-88*, vol. 24, no. 2, pp. 88–69, 2018.
- [10] J. N. Silva, J. Souza, Á. C. d. Oliveira, M. d. F. Tavares, L. C. Botega, et al., “Desenvolvimento de ontologia ciente de qualidade de informações para a melhoria de consciência situacional no domínio de gerenciamento de emergências,” 2017.
- [11] M. R. Endsley, “Design and evaluation for situation awareness enhancement,” in *Proceedings of the Human Factors Society annual meeting*, vol. 32, no. 2. SAGE Publications Sage CA: Los Angeles, CA, 1988, pp. 97–101.
- [12] R. R. Oleteo, “Percepção da qualidade da informação,” *Ciência da informação*, vol. 35, no. 1, pp. 57–62, 2006.
- [13] I.-G. Todoran, L. Lecornu, A. Khenchaf, and J.-M. L. Caillec, “A methodology to evaluate important dimensions of information quality in systems,” *Journal of Data and Information Quality (JDIQ)*, vol. 6, no. 2-3, pp. 1–23, 2015.
- [14] L. P. English, *Information quality applied: Best practices for improving business information, processes and systems*. Wiley Publishing, 2009.