Foto de Glenn Carstens-Peters na Unsplash

Desvendando o IMSC1: A Tecnologia por trás das legendas ocultas na TV 3.0

Por Vinícius Vasconcellos

Este artigo do Grupo de Trabalho de TV 3.0 da SET descreve a tecnologia IMSC1 - Internet Media Subtitles and Captions -, escolhida para a distribuição de legendas ocultas over-the-air e pela internet para o futuro padrão de TV 3.0.

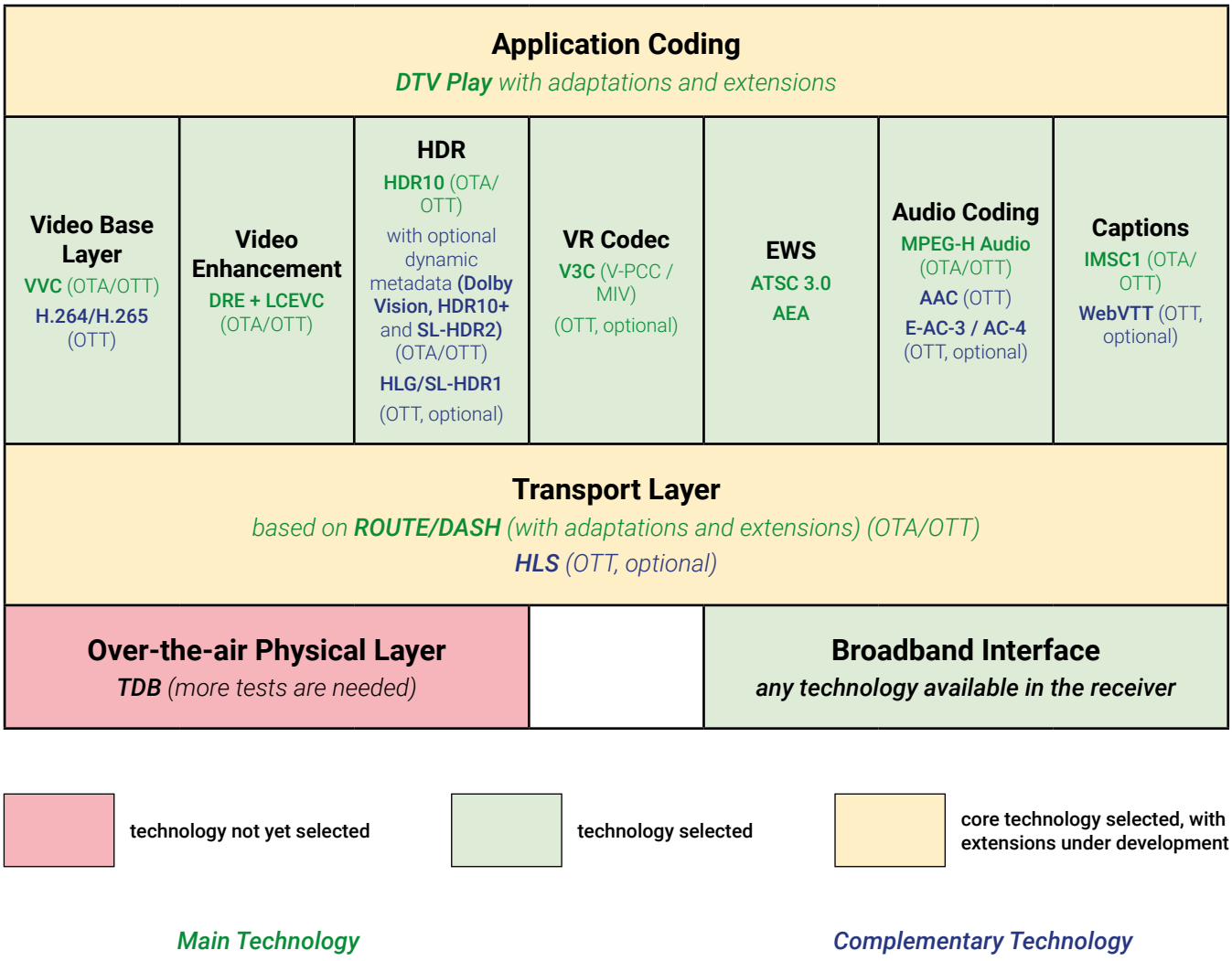
Introdução

Para o desenvolvimento do projeto "TV 3.0", o Fórum do Sistema Brasileiro de TV Digital Terrestre (SBTVD) tomou a iniciativa de abrir uma Chamada de Propostas, convidando organizações interessadas a submeterem suas tecnologias candidatas para integrar os diversos componentes e subcomponentes do sistema (detalhes disponíveis em: https://forumsbtvd.org.br/wp-content/uploads/2020/07/SBTVDTV-3-0-CfP.pdf).

Após um rigoroso processo de testes e avaliações, o padrão IMSC1 (***Internet Media Subtitles and Captions Profile 1***) foi escolhido na Fase 2 do Projeto da TV 3.0 [1] para ser adotado tanto para a distribuição ***over-the-air*** quanto via Internet, mantendo-se também um suporte opcional ao WebVTT (presente na especificação da TV 2.5). Esta decisão colocou o padrão IMSC1 à frente dos padrões ARIB-TTML (japonês) e AVS Captions (chinês), destacando-se como a escolha preferencial para o futuro da transmissão de conteúdo televisivo no país.

Com o avanço da era digital, observamos um aumento significativo no consumo de conteúdos de mídia, impulsionando a demanda por soluções avançadas que assegurem a acessibilidade e a interoperabilidade das legendas em vídeos. Neste contexto, o padrão IMSC1 emergiu como uma solução robusta e adaptável, projetada para satisfazer essas exigências com eficiência. Este artigo visa elucidar os princípios fundamentais do IMSC1, bem como seu embasamento teórico, enfatizando as aplicações práticas dessa tecnologia no cenário atual e os benefícios que proporcionará a sua implementação.

Figura 1 - Recomendações de tecnologias candidatas para a continuidade do Projeto TV 3.0 [1]



O que são legendas ocultas?

As legendas ocultas, também conhecidas como *closed captions* (CC) em inglês, representam métodos de inclusão textual de áudio em uma vasta gama de mídias, como programas de televisão, filmes, vídeos e outras mídias audiovisuais. Embora seu principal objetivo seja auxiliar pessoas com deficiência auditiva, elas se estendem a aplicações mais amplas, beneficiando também aqueles que se encontram em locais ruidosos, como o transporte público, ou estádios de futebol; aprendizes de idiomas; ou indivíduos que buscam uma concentração mais aprofundada no conteúdo assistido. Diferentemente das legendas tradicionais, as legendas ocultas podem ser ativadas ou desativadas pelo usuário quando necessário.

As legendas ocultas não apenas transcrevem diálogos, mas também incluem indicações de quem está falando e descrições de sons não verbais que são importantes para entender o contexto de uma cena, como "(buzina tocando)" ou "(música triste)". Elas são uma ferramenta importante de acessibilidade, garantindo que o conteúdo de mídia seja inclusivo e acessível a todos os públicos. Com o avanço da tecnologia e a crescente conscientização sobre a inclusão, o uso de legendas ocultas tem se expandido para além da televisão, abrangendo também serviços de streaming, vídeos online, jogos eletrônicos e outros formatos de mídia digital.

O que é IMSC1?

O IMSC1 [2] é um perfil do padrão IMSC, desenvolvido para fornecer legendas de alta qualidade e acessíveis para conteúdo de mídia online. O padrão IMSC foi elaborado pelo ***World Wide Web Consortium*** (W3C), a organização global que desenvolve padrões da Web, com o objetivo de estabelecer diretrizes comuns para a criação e distribuição de legendas, e define um conjunto específico de recursos e práticas recomendadas para garantir a interoperabilidade entre diferentes plataformas e dispositivos.

O IMSC1 foi desenvolvido como um formato candidato para atender os requerimentos do ***Interoperable Media Format*** (IMF), um esforço do SMPTE para desenvolver um único formato de arquivo de mídia mezanino destinado à transferência de mídia de estúdios para distribuidores, mas seu escopo se expandiu para o uso de arquivos também, no entanto não se destinava a ser o formato para a entrega pela internet aos clientes finais. A Netflix é uma notável apoiadora do IMF. O IMSC é derivado do CFF-TT, e incorpora aspectos do EBU-TT.

Na prática, o IMSC é uma linguagem de marcação utilizada para definir um texto cronometrado que adiciona legendas em mídias digitais. Ele define a estrutura do conteúdo da legenda em um documento XML, que consiste em uma série de elementos que podem ser utilizados para incluir ou agrupar diferentes partes do conteúdo da legenda para que apareça de uma determinada maneira ou em um determinado momento. Muitos destes elementos são semelhantes ou iguais aos recursos HTML [3].

O MDN Web Docs, o website oficial de Mozilla para desenvolvimento de padrões web e projetos de Mozilla, previamente conhecido como Mozilla Developer Network ou Mozilla Developer Center, disponibiliza um guia básico de IMSC em um ambiente onde é possível editar o HTML e CSS dinamicamente a partir de um documento XML IMSC - https://developer.mozilla.org/en-US/docs/Related/IMSC/Basics.

Cada documento IMSC é independente e combina informações de conteúdo, tempo, layout e estilo. O conteúdo do documento é estruturado usando ***tags*** semelhantes às usadas em HTML, como ***<body>, , <p>, *** e ***
. O tempo e o estilo são expressos usando atributos como início, fim, cor, ***tts:backgroundColor, ***tts:fontSize***, ***tts:fontFamily*** — esses são conceitos familiares para qualquer pessoa familiarizada com CSS (**NE**: CSS é chamado de linguagem ***Cascading Style Sheet*** e é usado para estilizar elementos escritos em uma linguagem de marcação como HTML. O CSS separa o conteúdo da representação visual).

O padrão IMSC1 [2] foi adotado pelo ATSC como a norma preferencial para legendagem, devido ao seu design que atende especificamente às necessidades de transmissão e distribuição digital por banda larga, conforme definido no documento normativo ATSC A/343 [4], que estabelece as especificações para a implementação de legendas em fluxos de mídia ROUTE/DASH e MMT (***MPEG Media Transport***), abrangendo aspectos como conteúdo, empacotamento e sincronização. Baseando-se

Figura 2: Documento IMSC básico renderizado – Ver este conteúdo escaneando o QR

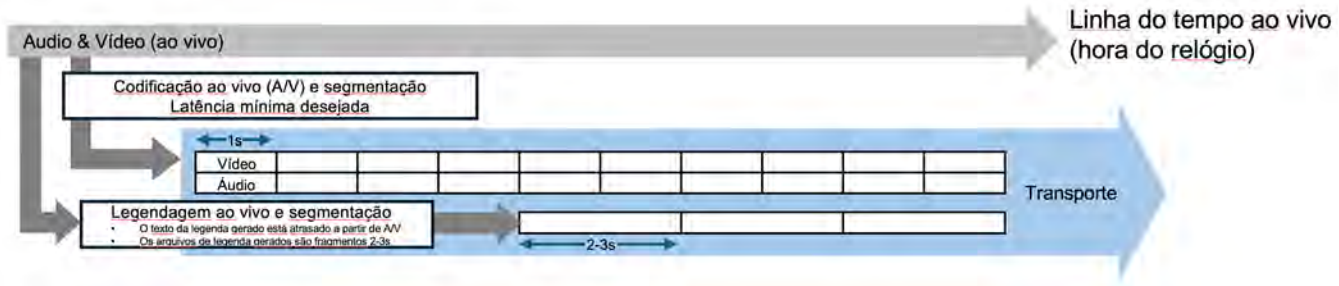
```
<tt xmlns="http://www.w3.org/ns/ttml" xml:lang="en">
  <body>
    <div>
      <p begin="1s" end="2s">Olá, a TV 3.0 está chegando!</p>
      <p begin="2s" end="3s">E o IMSC1 foi o padrão escolhido para legendagem.</p>
    </div>
  </body>
</tt>
```

na tecnologia ***SMPTE Timed Text*** (SMPTE-TT), especificada na norma SMPTE 2052-1 [5], o padrão assegura a compatibilidade com uma ampla gama de símbolos e idiomas globalmente, além de suportar a entrega de glifos de imagem. E sua aplicabilidade se estende por diversas plataformas de distribuição de mídia, incluindo serviços de radiodifusão e via Internet, destacando-se pela sua flexibilidade e capacidade de atender a requisitos avançados de legendagem.

A transmissão de legendas no padrão IMSC1 é realizada por meio de arquivos ISO BMFF, contendo documentos XML em conformidade com as especificações do IMSC1, como detalhado no ATSC A/343 [4]. Cada arquivo ISO BMFF abriga um conjunto de "texto cronometrado" que acompanha um conjunto correspondente de metadados de sinalização, garantindo uma integração eficaz com o conteúdo audiovisual.

Para a integração de legendas em conteúdos ao vivo ou pré-existentes, os segmentos de legendas em formato ISO BMFF (documentos IMSC1) são projetados com durações curtas, variando tipicamente de 1/2 até 3 segundos, ver **Figura 3**. Essa característica é crucial para permitir que os decodificadores sincronizem e apresentem as legendas em tempo real com o conteúdo de vídeo e áudio, otimizando a experiência do usuário ao minimizar possíveis atrasos na visualização das legendas.

Figura 3: Modelo de temporização de legenda ao vivo



Quais as diferenças entre o IMSC e o WebVTT?

O IMSC não tem relação com o WebVTT, que é outra forma de fazer legendas e legendas para a Web. As recomendações W3C para legendagem de mídia on-line, são desenvolvidas pelo ***Timed Text Working Group*** (TTWG) que mantêm atualizados os perfis de TTML (***Timed Text Markup Language***) e WebVTT (***Web Video Text Tracks***) com base na experiência de implementação e feedback de interoperabilidade, e na criação de mapeamentos semânticos entre essas linguagens. O grupo também desenvolve notas de uso e gerência registros para apoiar os implementadores [6].

O WebVTT (***Web Video Text Tracks***) [7] é um formato de código aberto para legendas e ***closed captions*** em vídeos online. É suportado por todos os principais navegadores web e players de mídia, como Chrome, Firefox, Safari, Edge, VLC e MPC-HC, enquanto o IMSC não é, no entanto, foi desenvolvido um ***polyfill*** IMSC, chamado ***imscJS*** [9], e que pode ser utilizado para renderizar os arquivos IMSC em HTML5. Da perspectiva de um desenvolvedor, o ***imscJS*** permite uma experiência consistente em todos os navegadores.

O IMSC oferece ainda funcionalidades adicionais para HDR e 3D estereoscópico que não estão disponíveis no WebVTT. Bem como, estilos adicionais como o ***tts:linePadding***, utilizado para controlar a quantidade de espaço vertical entre as linhas, e o ***tts:fillLineGap***, utilizado para distribuir uniformemente o texto dentro da área de exibição da legenda, ajustando a quantidade de espaço vazio entre as linhas e recursos, úteis para a visualização de legendas e legendas ocultas [8].

Figura 4: Extraído keynote "Jornada da TV 3.0", de Carolina Duca no SET Sudeste 2024/Reprodução

Benefícios do IMSC1

O IMSC1 apresenta múltiplos benefícios para sua implementação. Entre suas principais vantagens, destacam-se:

- **Eficiência na Distribuição**: O IMSC1 adota formatos de arquivo otimizados e compactos para a distribuição de legendas, o que contribui para a redução dos tempos de carregamento e minimiza o consumo de largura de banda. Essa eficiência é crucial em contextos onde os recursos de rede são limitados ou em que a entrega de conteúdo de alta definição é primordial. Ao otimizar a distribuição de legendas, o IMSC1 facilita uma experiência de visualização mais ágil e de maior qualidade para o usuário final. Plataformas de ***streaming*** de vídeo, como Netflix [10] e Apple [11], utilizam o IMSC1 para fornecer legendas em seus conteúdos. Isso garante uma experiência consistente e acessível para milhões de espectadores em todo o mundo.

- **Interoperabilidade**: Uma das mais significativas contribuições do IMSC1 é sua capacidade de garantir consistência operacional em diversas plataformas e dispositivos. Isso inclui navegadores de internet, aplicativos para dispositivos móveis, televisores conectados, entre outros. Tal característica assegura que os produtores de conteúdo possam criar legendas de forma unificada, promovendo sua distribuição para um espectro amplo de audiências sem a necessidade de ajustes adicionais. Essa universalidade não apenas simplifica o processo de produção, mas também amplia o alcance do conteúdo a diferentes públicos.

- **Acessibilidade**: Priorizando a inclusão, o IMSC1 se compromete a tornar as legendas acessíveis e compreensíveis para todos, incluindo pessoas com deficiências auditivas ou visuais. Para tanto, suporta a implementação de recursos como texto descritivo, ajustes na cor de fundo, variação no tamanho da fonte, entre outras funcionalidades que potencializam a legibilidade das legendas. Esta abordagem inclusiva não só cumpre com os requisitos legais de acessibilidade, mas também enriquece a experiência de visualização para um espectro mais amplo de usuários.

- **Compatibilidade com tecnologias futuras**: Diante do avanço acelerado da tecnologia e da constante atualização dos padrões web, o IMSC1 foi desenvolvido com foco na adaptabilidade e compatibilidade com inovações tecnológicas futuras. Essa característica assegura que as legendas criadas sob este padrão estejam preparadas para acompanhar os desenvolvimentos futuros, particularmente os avanços trazidos pela inteligência artificial em áreas como a conversão automática de voz para texto e a tradução para diferentes idiomas. Além disso, essa adaptabilidade facilita a integração das legendas em novas plataformas de mídia digital, garantindo que o conteúdo se mantenha relevante e acessível, mesmo diante das rápidas mudanças no cenário tecnológico.

A adoção do IMSC1 representa um passo significativo na direção de uma experiência de mídia mais inclusiva, eficiente e adaptável às necessidades futuras, solidificando sua posição como um padrão essencial no ecossistema de produção e distribuição de conteúdo digital.

Conclusões

Conforme apresentado, o padrão IMSC1 não é apenas um componente técnico essencial para a implementação da TV 3.0, mas também um marco importante na jornada em direção a uma sociedade mais inclusiva, onde o acesso à informação e ao entretenimento é garantido a todos, independentemente de suas capacidades auditivas ou visuais e desempenha um papel crucial na garantia de acessibilidade e interoperabilidade em legendas de mídia. O IMSC1 é uma ferramenta indispensável para a criação e distribuição de legendas acessíveis e de alta qualidade na televisão e na internet, e ressalta a importância da evolução contínua dos padrões de legendagem em resposta às transformações tecnológicas e às expectativas dos usuários. Ao adotar e promover padrões como o IMSC1, a TV 3.0 não apenas cumpre com suas responsabilidades sociais e legais de acessibilidade, mas também se posiciona na vanguarda da inovação tecnológica, preparada para os desafios e oportunidades que o futuro reserva.

Referências

SBTVD Forum: TV 3.0 Project. https://forumsbtvd.org.br/tv3_0/#panel-phase2

W3C: TML profiles for Internet media subtitles and captions. 2020. TTMLprofiles for Internet media subtitles and captions 1.2. https://www.w3.org/TR/ttml-imsc1.2/ Online; Accessed on August 30, 2023.

MDN Web Docs: IMSC Basics https://developer.mozilla.org/en-US/docs/Related/IMSC/Basics

ATSC Standard: Captions and Subtitles, with Amendments No. 1 and No. 2, 2021, https://www.atsc.org/atsc-documents/a3432017-atsc-standard-captions-subtitles/.

Technology Committee 24TB. 2013. ST 2052-1:2013 - SMPTE Standard - Timed Text Format (SMPTE-TT). ST 2052-1:2013 (June 2013), 1–18. https://doi.org/10.5594/SMPTE.ST2052-1.2013

W3C: W3C Wiki Timed Text Working Group. https://www.w3.org/wiki/TimedText

W3C: WebVTT: The Web Video Text Track Format https://www.w3.org/TR/webvtt1/

MDN Web Docs: IMSC: subtitles and captioning for the Web https://developer.mozilla.org/en-US/docs/Related/IMSC#differences_between_imsc_and_webvtt

imscJS: https://github.com/sandflow/imscJS

Netflix: Netflix IMSC 1.1 Text Profile https://partnerhelp.netflixstudios.com/hc/en-us/articles/360053755033-Netflix-IMSC-1-1-Text-Profile

Apple: HTTP Live Streaming (HLS) authoring specification for Apple devices https://developer.apple.com/documentation/http-live-streaming/hls-authoring-specification-for-apple-devices

Vinícius Vasconcellos é consultor Sênior na IBM e Conselheiro da SET. Na IBM trabalha com foco na aplicação de novas tecnologias como Inteligência Artificial, Cloud, EDGE Computing, Big Data e Analytics na indústria de Telecom, Mídia e Entretenimento. Com sólida experiência em telecomunicações e broadcast, atuou como especialista e executivo na Claro e no Grupo RBS. É formado em Engenharia Elétrica com especialização em Administração pela UFRGS e possui MBA em Gestão Executiva pela Fundação Dom Cabral.

Contato: ***vasv@br.ibm.com***