



PDF to XML Conversion

CASE HIGHLIGHTS

The Client

Company: Open access medical journal, focusing on health innovation for global community

Location: Australia

Industry: Health Care

The Objective:

Converting Open Access Medical Journal from pdf files to XML format

The Solution:

PDF to XML conversion process through Adobe PDF, PubMed style checker & PubMed Article Previewer, helped clients save 30-40% on costs.

The Challenges:

- Non unique font or font size where both words and numerical meaning delimiters were in different font size.
- PDF source document contained peculiar typesetting errors. Few additions were shown in regular font, but others were in different fonts.
- Second stage of the conversion process required manual fixes to be carried out
- The font and font size attributes in the XML file needed detailed inspection in order to detect and mitigate the whitespace characters to avoid erroneous typeface information that could cause clutter across the transitions in the XML process.

Recent years have brought about evolutionary changes in efforts to digitize medical data, however; integrating data from various providers, and across domains, remains a challenge till date. Simultaneously technologies have also evolved towards resolving these challenges. However, the progress is slow, and this is due to the scarcity of domain specific software tools that address specific challenges.

One of the leading periodicals, printers and publisher approached us for conversion of Open Access Medical Journal from pdf files to XML format. The process also warranted data collection in XML editor, implementing tagging guidelines approved by PubMed Central and validation of XML files against the XML schema.

Technology / Software used

Adobe PDF, PubMed style checker & PubMed Article Previewer were used after comprehending client requirements for converting data from PDF format to XML format.

The Solution

The process was streamlined, starting with collection of Input PDF files from various sites suggested by client.

- Nomenclature changed based on PubMed guidelines, followed with XML conversion carried out while inserting XML tags as per XML schema approved by PubMed Central.
- Stringent Quality Control parameters were followed to validate XML files against the XML schema.
- Style Checker QC was opted for checking the final XML with online PubMed Style checker to confirm whether an XML file conforms to PMC Style as defined in the PMC XML Tagging Guidelines.
- Files then were viewed on PubMed Article Previewer giving you that feel and look of an article as it would appear in PubMed Central & finally the files were uploaded on client's FTP server.

Benefits

- Effective and Economical way to publish data on web, led to client satisfaction
- Entire conversion process enhanced the efficiency of data access
- Client was able to retain complete PDF document hyperlinks
- Enabled extraction of text without images; if required
- Entire content allowed editing and formatting the converted output for redistribution

Contact Us

Phone : +91-794-000-3000
+91-793-014-3000
Fax : +91-794-000-3202
E-Mail : info@hitechos.com

About Hi-Tech Outsourcing Services

Hi-Tech BPO, ISO 9001:2008 Certified service providers, an India based organization providing comprehensive business process outsourcing services to 3,100 plus, 80% recurring, clients across USA, Europe, Australia & Middle-East. The company offers a full range of BPO services across industries like Healthcare, Medical, Banking, Finance, Media and Entertainment, Hospitality and Travel - to name a few. Known as an innovative and trusted partner for Business Process Outsourcing (BPO) Services, their service portfolio includes data entry, data processing, data conversion, web research, scanning and indexing, transcription services and much more. Rated as a high performance company, they implement a multilayered quality check for all projects and ensure that qualitative, time bound and cost effective services are consistently delivered.

Hi-Tech Outsourcing specializes in delivering data conversion services for ePublication, Inventory & Media transcription, Legal, Medical, Financial, services in a qualitative & time bound manner, to individuals, academicians, attorneys, medical practitioners, & companies across the globe. Be it specialized or general transcription; they help you transcribe all digital voice, audio and video file formats, microcassette, tapes, DVD, CD, MP3, MP4 & WAV conveniently. Their consistent efforts at building long-term relationships with clients, backed by commitment to deliver on-time and qualitative services, have been pivotal to the consistent growth above market standards.

Headquarters Ahmedabad:

Hi-Tech House,
Near Gurukul Tower, Gurukul,
Ahmedabad - 380052 Gujarat,

Ahmedabad:

3rd Floor, Chanakya Plaza,
Opp. Swagat I, New C.G. Road,
Chandkheda, Ahmedabad - 382424

USA (California):

547 Moreland Way
Santa Clara, CA 95054
Phone: +1-408-338-8047
E-Mail: sales@hitechos.com

Europe (Denmark):

Scandinavia
Indslevvej 26, 5580,
Nr. Aaby, Denmark

Ahmedabad:

5th & 9th Floor, Commerce House - 4,
100 Feet Road, Prahladnagar,
Ahmedabad - 380015

Cochin:

43-A, E Block, 3rd Floor,
Cochin Special Economic Zone (CSEZ),
Kakkanad, Cochin - 682037
Kerala, India

USA (New York):

50 Riverside Drive,
New York, NY 10024,USA

Europe (Benelux):

Laathof 24 4851 ET,
Ulvenhout, Netherlands