

Digitalización de documentos para su utilización en una biblioteca virtual

Bargiela Roberto, Dapozo Gladys N., Bogado, Verónica S.

Departamento de Informática. Facultad de Ciencias Exactas y Naturales y Agrimensura.
Universidad Nacional del Nordeste.

roberto.bargiela@yahoo.com.ar, gndapozo@exa.unne.edu.ar, vro.s.bg@gmail.com,

Resumen

En el ambiente educativo las nuevas tecnologías se han vuelto una herramienta indispensable, han permitido traspasar barreras de espacio y tiempo, permitiendo capacitar a personas inmersas en diferentes culturas. En este contexto la digitalización es un paso importante e indispensable para la publicación de documentos con la información necesaria para el fin mencionado. La metodología propuesta pretende normalizar el proceso por el cual se obtienen los documentos a ser publicados dentro de una Biblioteca Virtual.

Palabras clave: educación a distancia, plataforma e-learning, biblioteca virtual, digitalización de documentos.

Introducción

En el ambiente educativo las nuevas tecnologías se han vuelto una herramienta indispensable, han permitido traspasar barreras de espacio y tiempo, permitiendo capacitar a personas inmersas en diferentes culturas, bajo condiciones sociales y económicas diferentes. Sin embargo, esto se ha logrado gracias a los aportes de varias disciplinas, donde se mezclan aspectos pedagógicos, informáticos y de comunicación.

La educación a distancia es una estrategia educativa basada en la aplicación de la tecnología al aprendizaje sin limitación de lugar, tiempo, ocupación o edad de los estudiantes. Implica nuevos roles para los alumnos y los profesores, nuevas actitudes y nuevos enfoques metodológicos [1].

Según D. Rowntree [1], la formación a distancia es aquel sistema de enseñanza en el que el estudiante realiza la mayor parte de su aprendizaje por medio de materiales didácticos previamente preparados, con escaso contacto directo con los profesores, pudiendo o no tener

un contacto ocasional con otros estudiantes. En este concepto se destaca como elemento primordial el material (documentos de diferentes tipos) pedagógico, la información a la cual el alumno debe tener acceso para el proceso de aprendizaje.

El término educación a distancia cubre un amplio espectro de diversas formas de estudio y estrategias educativas, que tienen en común que ellas no se cumplen mediante contigüidad física continua, de profesores y alumnos en locales especiales para fines educativos; esta nueva forma educativa incluye todos los métodos de enseñanza en los que debido a la separación existente entre alumnos y profesores, las fases interactiva y preactiva de enseñanza son conducidas mediante la palabra impresa, y/o elementos mecánicos y electrónicos [1].

La mayoría de los autores destacan varios ítems que caracterizan a la educación a distancia siendo algunos de los más importantes los que se listan a continuación.

- Distancia entre profesor y alumno
- Preproducción: el curso suele estar preproducido, en material impreso o en otro medio (digital, visual o auditivo)
- Medios técnicos: material impreso y/o digital (texto, imágenes, sonido, video) con el contenido del curso.
- Autoinstructiva: el material debe estar preparado para que el alumno puede aprender sin necesidad del profesor.
- Estudio individual: el alumno debe aprender en forma individual, raras veces hay interacción con otros alumnos.
- Comunicación: debe existir una comunicación (sincrónica, en tiempo real, y asincrónica, diferida) entre el alumno y la institución u profesor.
- Gran número de estudiantes y mínimos gastos.

- Métodos industriales: el proceso de enseñanza aprendizaje al aplicarse en forma masiva debe ser industrializado, es decir, se deben desarrollar métodos de planeamiento, conducción y evaluación del aprendizaje automatizados y racionalizados.

Los medios tradicionales de comunicación (correo postal, radio, TV) junto con el material impreso, audio- visual, han sido los orígenes de la educación a distancia. Hoy en día, este concepto de educación viene asociado a lo que se conoce como nuevas tecnologías de la información y comunicación. Los entornos virtuales de aprendizaje, nuevos escenarios, e-learning, teleformación, educación virtual, aprendizaje online son un conjunto de tecnologías que se emplean para la implementación del proceso de enseñanza-aprendizaje a distancia. Dichas tecnologías brindan nuevas y diferentes formas de acceso y utilización de la información y nuevas formas de comunicación (sincrónica y asincrónica) [2].

La evolución de las nuevas tecnologías en el contexto de enseñanza-aprendizaje está sujeta a cómo se insertan éstas en la práctica educativa y en el entorno socio-económico y cultural. Además, requieren la capacitación del formador (profesor) para que planea y guíe el procesamiento de la información para el aprendizaje y evalúe los resultados. Y preparación del destinatario (alumno) para que aprenda a elegir, analizar, evaluar la información y cumpla con los objetivos planteados, todo en el marco de las TIC. Las plataformas educativas implican un sistema complejo donde interactúan diversas culturas, trabajos colaborativos y tecnologías de diferentes tipos, construyendo redes de alumnos, docentes e investigadores. Las plataformas educativas o herramientas de e-learning abren un nuevo camino en la educación, creando aulas virtuales en la que los alumnos asisten en horarios diferentes y desde lugares diferentes, inclusive los alumnos tienen edades, expectativas y necesidades distintas, entre otras cosas.

Las plataformas educativas se hacen cada vez más populares, no sólo para estudiantes a distancia sino también para la enseñanza tradicional como herramienta complementaria, provocando que dichas plataformas crezcan en cuanto a posibilidades de brindar nuevos servicios además del puro hecho de “asistir a una clase virtual”, también es necesario y útil tener el material didáctico, los libros sugeridos por las cátedras y hasta inclusive apuntes redactados por los propios profesores y alumnos.

Como se ha visto el aspecto más importante de la educación a distancia es el material que se emplea para el contenido y soporte.

Lamarca Lapuente [4] afirma que en el terreno de la ciencia documental, la mayor parte de los autores coinciden en que un **documento es, esencialmente, información, la materialización de un mensaje o el soporte de una información**. Un documento, no es ni más ni menos, que un soporte para transferir información.

Según la autora, en la ciencia de la documentación, el documento es a la vez medio y mensaje de información y conocimiento. De esta manera, el documento se caracteriza por una triple dimensión: el soporte físico o material, el mensaje informativo y la posibilidad de transmisión o difusión de este conocimiento. Esta triple dimensión que caracteriza al documento ha servido, a su vez, para establecer una tipología de los documentos en la que los estudiosos del campo de la documentación coinciden, más o menos, y que se puede establecer de la siguiente forma:

- Según el soporte material:
 - papel (libros, revistas, folletos, etc.)
 - material químico (películas)
 - material magnético (cintas de vídeo, disquete de ordenador, casetes, etc.)
 - soportes ópticos (CD-ROM, DVD, video-discos, etc.)
 - material plástico (microformas: microfichas)
- Por el nivel y rigor del contenido:
 - científicos
 - técnico-profesionales

- culturales-divulgativos
- sociales
- Por la transformación del contenido:
 - primarios: originales
 - secundarios: hacen referencia a los primarios (bibliografías, catálogos, índices, sumarios, etc.)
 - terciarios: tienen estructura formal de secundarios, pero contenido primario (diccionarios, enciclopedias, léxicos, tesauros, etc.)
 - mixtos: tienen elementos primarios y secundarios (bibliografías comentadas, resúmenes o abstracts, etc.)
- Según la posibilidad de transmisión o difusión:
 - social:
 - públicos
 - reservados
 - inéditos
 - personales
 - temporal
 - periódicos
 - no periódicos

La era digital ha producido otro tipo de soporte para la información, debido al lenguaje con que trabajan las computadoras y la forma en que se transmite dicha información a través de las redes.

Un **documento electrónico** es un documento contenido en un soporte electrónico que, para su visualización requiere una pantalla textual, una pantalla gráfica, y/o unos dispositivos de emisión de audio, vídeo, etc; según el tipo de información que contenga. En algunos casos también se precisa la mediación de un ordenador (cuando la información está digitalizada), en otros no (si se trata de información analógica. A su vez, un documento electrónico puede ser analógico o digital, la diferencia entre estos es la forma en que están codificada la información de los mismos. En un documento digital la información está codificada en bits (binary digits) y para leer, visualizar o grabar la información se precisa de un dispositivo que transmita o grabe información codificada en bits. Al representarse digitalmente, los datos de entrada son convertidos en dígitos (0,1)

inteligibles para una computadora y no para los sentidos humanos y a la salida otro dispositivo los convertirá en señales analógicas, inteligibles para los sentidos humanos)[4].

| | | |
|----------------------------|--------------------------------|----------------------|
| Documentos impresos | Documentos electrónicos | |
| | Documentos analógicos | Documentos digitales |

Cuadro 1: Dicotomía entre documento ipresos y electrónicos, a su vez dentro de los documentos electrónicos; entre documentos analógicos y digitales

Codina [3][4] en su libro resume las características que definen a los documentos analógicos versus documentos digitales.

| Parámetro | Documento analógico | Documento digital |
|---------------------|---|--|
| Grabación | Diversos medios para grabar cada una de las diversas morfologías de la información | Un único medio para grabar cada una de las diversas morfologías de la información. |
| Codificación | Cada morfología de la información requiere su codificación específica, que cambia, además, según el soporte. | Todas las morfologías de la información se codifican como cadenas de ceros y unos, sea cual sea el soporte físico: chip de memoria RAM, disco duro, disco CD-ROM, etc. |
| Formas de lectura | Algunos documentos analógicos se pueden leer simplemente acercándolos a la luz, por ejemplo, los documentos impresos en papel. Es decir, poseen impacto visual directo. | No tienen impacto visual directo. Siempre necesitan un aparato de lectura. |
| Aparatos de lectura | Algunos documentos analógicos requieren un aparato de lectura, por ejemplo, las grabaciones de vídeo requieren un | Todos los documentos digitales requieren un aparato de lectura, más un ordenador. |

| | | |
|-------------------------|--|---|
| | magnetoscopio y un monitor de TV. | |
| Soporte de impresión | Cada morfología de la información se imprime en un soporte propio: papel, pantalla de cine, monitor de televisión, etc. | Toda la información se imprime en un único soporte: un monitor de ordenador. |
| Grado de interactividad | La interactividad entre el documento y el lector es muy reducida, y a menudo limitada a una simple conmutación, por ejemplo: apagar y encender un aparato, cambiar de canal, abandonar la lectura y volver a retomarla, etc. | Teóricamente, no tiene límites |
| Circulación | La circulación de los documentos implica necesariamente el desplazamiento de objetos físicos: libros, discos, cintas, etc. ("movimiento de átomos") | La circulación de los documentos no implica, necesariamente, desplazamiento de objetos físicos, sino únicamente desplazamientos de energía ("movimientos de bits"). |

Cuadro 2: Características de documentos analógicos y digitales. Comparativa

Las bibliotecas juegan un papel muy importante en cuanto a la organización, clasificación y búsqueda de material bibliográfico. Formalmente, una biblioteca es una colección de documentos, debidamente organizada y que sirve para cubrir las demandas generales y específicas de información [5].

El estudio a distancia implica poder tener el material disponible para que el alumno tenga acceso al mismo en forma rápida y sencilla.

Una nueva tecnología permite lograr estos propósitos mediante la aplicación del concepto de biblioteca tradicional al ambiente virtual, integrando tecnología informática y de comunicación en el ambiente de las redes, principalmente Internet. Dicha tecnología se conoce como Biblioteca Virtual.

Una Biblioteca Virtual, Digital o Electrónica es una colección de documentos digitales debidamente organizada con criterios biblioteconómicos, siguiendo estándares para los formatos digitales, con el fin de ponerla a disposición de los usuarios.

La implementación de una biblioteca dentro de un ambiente virtual requiere al igual que una biblioteca tradicional, una metodología adaptada a la naturaleza del requerimiento. Las etapas básicamente se corresponden con la labor habitual de cualquier biblioteca tradicional, la única diferencia entre una biblioteca tradicional y una virtual, son las técnicas o métodos y herramientas que utilizan para llevar a cabo cada etapa.

Existen diferentes metodologías que se han utilizado en los últimos tiempos para crear bibliotecas virtuales alrededor del mundo, siendo las etapas básicas: Selección, Catalogación, Clasificación, Indización y Digitalización.

Debido a la naturaleza virtual, resulta interesante el estudio de la etapa de digitalización. Dicha fase puede realizarse inmediatamente después de haber seleccionado al documento o en cualquier otro momento, eso dependerá de las necesidades, prioridades y otras consideraciones que posea cada biblioteca. Básicamente esta etapa consiste en el proceso de conversión o creación de un documento digital.

El proceso de digitalización varía según el contexto de aplicación. Existen varias metodologías para la publicación de documentos que definen los pasos para digitalizar documentos de acuerdo a sus objetivos. Los pasos a seguir quedan

1 Biblioteca Virtual, Digital o Electrónica suelen tomar significados similares en algunos casos y disímiles en otros, por ello en el presente trabajo se toman los tres nombres como sinónimos.

determinados en función del tipo de documento que se desea digitalizar.

Al proceso de digitalización se someten, generalmente, documentos impresos o manuscritos o analógicos, y dependiendo el caso a documentos audiovisuales, como fotos, películas o grabaciones de audio. Cada tipo de documento tendrá que pasar por métodos distintos de digitalización y el resultado de cada uno, también será distinto pero en términos generales, estarán en formato digital. Existen formatos digitales estándares ampliamente utilizados para cada uno de los tipos de documentos. Para documentos escritos suelen utilizarse los formatos PDF, HTML, Postscript, Documento de Word, Documento OpenOffice.org, etc. Para los documentos visuales son ampliamente utilizados los formatos AVI, MPEG, VCD o DVD y recientemente MP4. Para documentos de sonido se utilizan los formatos WAV y MP3. Por último, para las imágenes se utiliza el formato JPEG y en menor medida los formatos GIF y PNG, entre otros.

Cada formato tiene sus ventajas y desventajas, como así también, su adaptabilidad para diferentes situaciones. La decisión por uno o por otro formato se considera en función de las características del documento a digitalizar o de las necesidades y limitaciones de la biblioteca. La mayoría de estos formatos poseen una característica altamente explotada por las bibliotecas y es el hecho de que soportan metadatos, estos son datos incrustados dentro del documento digital que permite la realización de búsquedas y el alta de los mismos de manera más sencilla.

El estado del arte ha dejado en descubierto la importancia que tiene el proceso de digitalización en la publicación de documentos en forma electrónica dentro de un ambiente educativo, el cual es responsable de formatear el material imprescindible para cumplir con los objetivos pedagógicos planteados para el proceso enseñanza-aprendizaje a distancia.

El presente trabajo forma parte de un proyecto mayor cuyo objetivo es el desarrollo de una biblioteca virtual dentro de una plataforma de e-learning open

source, orientada especialmente a su utilización en instituciones de educación superior. En este subproyecto se pretende definir una metodología de digitalización adecuada a los requerimientos y restricciones especificados para el proyecto macro.

Metodología

El estudio de los diferentes procedimientos para digitalizar documentos ha proporcionado las bases para poder definir una metodología adaptada a los requisitos definidos. Esta metodología constituirá una etapa dentro del procedimiento necesario para implementar una biblioteca virtual como componente de una herramienta e-learning open source.

Digitalización de Documentos

En un proyecto específico de enseñanza a distancia, el medio en que se encuentra el material didáctico puede variar, pero debido a las características de las plataformas educativas, la forma en que se configura el proceso de enseñanza-aprendizaje y la naturaleza de las bibliotecas virtuales, la información relacionada con el tema de estudio, generalmente está contenida en papel (libros, revistas, folletos, etc.) y soportes ópticos (CD-ROM, DVD, video-discos, etc.). Estas características inherentes al documento a digitalizar restringirán el procedimiento a seguir al igual que el tipo de formato digital a producir.

Cabe aclarar que pueden existir documentos digitales que por conveniencia económica o práctica u otra razón se convertirán a otro formato digital, estos documentos también atravesarán el proceso de digitalización.

Proceso de Digitalización

El proceso de digitalización consta esencialmente de los siguientes pasos:

1. Determinar el medio en el que el documento se encuentra soportado. A los efectos prácticos se consideran susceptibles de ser digitalizados los documentos en los siguientes medios.
 - a) papel: (libros, revistas, folletos, etc.)

- b) material químico (películas)
 - c) material magnético (cintas de vídeo, disquete de ordenador, cassetes, etc.)
 - d) soportes ópticos (CD-ROM, DVD, video-discos, etc.)
2. Según el medio determinado en el paso anterior, se procede a digitalizarlo o a convertirlo a otro formato digital distinto del original.
- a) Definir el hardware/software: a utilizar según el medio que soporta el documento o el formato electrónico del mismo.
 - b) Determinar el formato digital final: para el presente trabajo se define en primera instancia como formatos básicos los estándares, los formatos más usados a nivel usuario.
 - c) Proceder a la digitalización propiamente dicha. Este paso depende de los factores determinados anteriormente (medio, formato, hardware y software correspondientes)
3. Determinado el formato digital final y digitalizado el documento original, se completa el documento digital con los metadatos necesarios para su clasificación, catalogación e indización.

En algunos casos es posible que en el paso 2 de la metodología propuesta surja un documento digital intermedio, es decir que estará en un formato digital que servirá de origen para el formato digital final.

Para cada uno de los medios se utilizarán dispositivos y/o software diferentes para llevar a cabo la digitalización. Para todos los casos se utilizará una PC con el hardware mínimo.

A continuación se listan los requisitos mínimos para la digitalización.

| Medio | Hardware | Software |
|--------------------|--|---|
| Papel | Scanner o multifunción | Reconocedor óptico de caracteres, procesador de textos |
| Material Magnético | Vidreoreproductor, TV, lector de disco de 3"1/2 y/o 5"1/4, equipo de audio con | Edición de vídeo, edición y mezcla de audio, Grabador de CD/DVD |

| | | |
|------------------|---|---|
| | cassetera y conexión a PC, Placa de vídeo con salida a TV, captura de vídeo y/o sintonizadora de TV. Placa de sonido con capacidades para conectarse a equipos de audio. Lectograbadora de CD/DVD | |
| Soportes Ópticos | Lectograbadora de CD/DVD. Placa de sonido con capacidades para conectarse a equipos de audio. | edición y mezcla de audio, Grabador de CD/DVD |

Cuadro 3: Requisitos para digitalizar

En este punto (paso 2 de la metodología propuesta) es necesario determinar el formato digital final que se desea producir. Para ello, se debe tener en cuenta la morfología del documento. Los formatos digitales ampliamente utilizados en el mundo y los determinados para la implementar en el trabajo propuesto son los siguientes:

- a) Texto (con o sin imagen): pdf
- b) Imagen: el formato en este caso depende de la calidad de la imagen que se requiera obtener y de las características de la imagen original, pero fundamentalmente los elementos que se deben considerar son la resolución, el tamaño físico en disco del archivo y la profundidad de colores. Los formatos que se considerados hasta el momento son: jpeg, gif, png.
- c) Audio: al igual que el ítem anterior se deben considerar un conjunto de atributos que definen la calidad del sonido. Los formatos más comunes que se tomarán como base son: mp3, wav.
- d) Vídeo: incluye aspectos de imagen y audio. Por lo tanto la calidad queda determinada en función de lo anterior. Los formatos que se utilizarán: avi, mpg, vcd, dvd, wmv.

En los formatos digitales para documentos audiovisuales hay que tener en cuenta la

calidad del documento ya que si se elige un formato de baja capacidad para representar la información es posible que se produzca pérdida de información lo cual se traduce en una disminución de la calidad; por ejemplo pérdida de colores, pérdida de calidad de audio, disminución en la nitidez de la imagen, etc.

Quizás el documento de origen que requiere más atención sean los documentos impresos como libros, revistas, artículos, etc.

Conclusiones

La digitalización de documentos es parte fundamental de toda biblioteca virtual y la documentación y material didáctico es parte indispensable en el proceso de enseñanza-aprendizaje a distancia.

Respecto a la metodología de digitalización, se puede concluir que es apropiada al contexto en que se desarrollará el trabajo. Sin embargo, resulta imprescindible que la misma sea revisada y adaptada a las nuevas tecnologías en cuanto a hardware y software.

En cuanto a los formatos, se ha determinado que debido a la naturaleza de la aplicación es conveniente utilizar principalmente los mencionados en la metodología, teniendo como ventaja que son adaptables, ampliamente empleados por los diferentes tipos de usuarios y existen una gran variedad de aplicaciones que manipulan dichos formatos.

En trabajos futuros, se utilizará esta metodología para la etapa de digitalización de los documentos que serán colocados en una biblioteca virtual, dentro de una plataforma educativa con el fin de que el alumno tenga la información necesaria para cumplir los objetivos de formación propuestos. La biblioteca permitirá almacenar el material de forma más eficiente y acceder al mismo de forma más rápida mediante la aplicación de algoritmos de búsqueda y clasificación.

Referencias

[1] L. Gracia Aretio. "Un concepto integrador de educación a distancia". Publicado en forma impresa en Radios y comunicación de adultos. Boletín nº 17, 1991.

En:<http://www.uned.es/catedraunesco-ad/articulos/1991/un%20concepto%20integrador%20de%20enseñanza%20a%20distancia.pdf>

[2] A. B. Tedesco. "Educación a distancia: entre los saberes y las prácticas". Septiembre de 2.005.

En: <http://weblog.educ.ar/educacion-tics/archives/006001.php>

[3] L. Codina. El libro digital y la WWW. Madrid, Tauro Ediciones, 2000.

[4] Ma. J. Lamarca Lapuente, Tesis Doctoral Hipertexto: El nuevo concepto de documento en la cultura de la imagen. Universidad Complutense de Madrid. Facultad de Ciencias de la Información. Dpto. de Biblioteconomía y Documentación. Doctorado: Fundamentos, Metodología y Aplicaciones de las Tecnologías Documentales y Procesamiento de la Información, Fecha de Actualización: 03/04/2007. Consultado el 2 de Abril de 2007.

[5] F. Herrera Diez. "RDIB (Recursos De Interés Bibliotecario)". Última consulta 2 de abril de 2.007.

En: <http://www.felixherreradiez.com/glos.htm>