

Universidad Autónoma Metropolitana

Unidad Azcapotzalco
División de Ciencias y Artes para el Diseño
Doctorado en Diseño y Visualización de la Información

**Minería de textos aplicada a los programas de estudio
de la carrera de DCG de CyAD en la UAM-A**

Francisco E. Torres García
franciscoetorresg@gmail.com
Twitter @ftorres2706
<http://bit.ly/TorresGF>

Ponencia presentada para el
1er. Coloquio de Educación para el Diseño en la Sociedad 5.0
Abril 2018

Minería de textos aplicada a los programas de estudio de DCG

Resumen

Los Programas de Estudio (PE) de Diseño de la Comunicación Gráfica (DCG) encierran información que se extrae mediante procesos de minería de textos, que revelan una orientación académica tradicional apoyada en la exposición presencial de contenidos, también se nota un sesgo hacia la mitad inferior de la Taxonomía de Bloom de objetivos de aprendizaje.

La composición y extensión de los apartados se descubre desigual; no se presenta un enfoque por competencias (el término aparece solamente en una Unidad de Enseñanza Aprendizaje (UEA)¹), enfatizándose las competencias llamadas duras (saber y saber hacer) y carencia en las blandas (saber ser y saber convivir). La bibliografía es notoriamente antigua (32% con más de 20 años, 74% con diez años o más), su redacción y citación carecen de uniformidad.

Las referencias electrónicas son genéricas y escasas y, salvo en la línea de estudio de Tecnología, que abarca siete UEA, los términos asociados a las Tecnologías de la Información y las Comunicaciones (TIC) están apenas presentes, el vocablo “Internet” aparece solamente en cuatro de las 99 UEA, siendo tres de ellas de la línea de Historia y la otra es creación de páginas web².

En conclusión, los hallazgos son inquietantes, es necesario discutir la manera en que los PE puedan actualizarse rápida y frecuentemente a fin de servir mejor a los alumnos.

- *Palabras clave: DCG, Programas de estudio, Minería de textos, Taxonomía de Bloom.*

Minería de textos

La *minería de datos*, y su subconjunto, la de textos (enfocada en elementos alfabéticos y relaciones semánticas), es un procedimiento metódico, posibilitado por tecnologías digitales y relacionado con técnicas de acceso y organización de información, hacia

detectar información procesable de conjuntos de datos. Utiliza análisis matemático para deducir patrones y tendencias que existen en ellos. Normalmente, estos patrones no se pueden detectar mediante exploración tradicional porque las relaciones son complejas o hay demasiados datos (Microsoft, 2017). O sea

...el descubrimiento de patrones interesantes y nuevos conocimientos en una colección de textos, es decir, [...] el proceso encargado del descubrimiento de conocimientos que no existían explícitamente en ningún texto de la colección, pero que surgen de relacionar el contenido de varios de ellos (Hearst, 1999; Kodratoff, 1999) citados en (Montes-y-Gómez, 2003, pág. 4).

¹ Gestión del Diseño II (Desarrollo Profesional) [1402020].

² Historia del Diseño Gráfico I (Arte y Diseño Gráfico) [420044], Historia del Diseño Gráfico II (Historia del Diseño Gráfico) [1420050], Cultura y Diseño II [1400030] y Tecnología para el DG V (Páginas Web) [1402057].

Con esa técnica se examina un contenedor de información en formato tabular, se concentran y analizan los vocablos y frases que lo componen, se busca hacer inferencias, descubrir patrones en los contenidos y hacer interpretaciones o predicciones con base en esos datos a través de la presencia y frecuencia o ausencia de ciertos elementos.

En este estudio se recurre a una minería de textos básica, por ser un volumen reducido de textos; las herramientas usadas son el procesador de texto Microsoft Word, la hoja de cálculo Microsoft Excel, la aplicación PureText 4.0, de Steve Miller (2018), el programa Primitive Word Counter versión 2.0 (Primitivezone, 2013) y la página web wordclouds.com (Zygomatic, 2015) que facilita generar nubes de texto.

Análisis Bloom

Una forma de analizar los programas de estudio es mapeando los verbos que en ellos aparecen con respecto a la Taxonomía de Bloom para la era digital (Heer, 2012), un modelo cognitivo jerárquico enfocado a categorizar los objetivos educativos en tres dominios psicológicos: cognitivo, afectivo y psicomotor.

En el dominio cognitivo, el paradigma muestra seis niveles, de izquierda a derecha desde lo elemental hasta lo complejo, resaltando los verbos y acciones de cada nivel:

Orden bajo						Orden alto					
1	2	3	4	5	6	1	2	3	4	5	6
Recordar	Entender	Aplicar	Analizar	Evaluuar	Crear	<i>Recordar información fechas, lugares, hechos, etc.</i>	<i>Dar significado a información para formar un acervo</i>	<i>Utilizar conocimientos a fin de solucionar problemas</i>	<i>Dividir el todo en elementos a fin de entenderlo</i>	<i>Comparar situaciones con la experiencia y emitir juicios de valor</i>	<i>Integrar diversas partes y construir un todo</i>
<i>Citar nombrar enlistar recordar...</i>	<i>Deducir explicar distinguir identificar...</i>	<i>Manejar graficar preparar utilizar...</i>	<i>Dividir analizar comparar clasificar...</i>	<i>Evaluuar criticar clasificar justificar...</i>	<i>Crear idear diseñar elaborar...</i>						

Tabla 1. Taxonomía de Bloom para la era digital, con verbos emblemáticos. Fuente: Heer (2012), traducida y resumida por Torres, F.

Programas de estudio

La carrera de DCG de la UAM-A contempla 99 PE para sus UEA, publicados por la Coordinación de Docencia (2016). Los 15 campos principales que componen cada programa son: trimestre, clave UEA (siete dígitos), nombre, créditos, tipo (obligatoria u optativa), Tronco (General, Básico, Profesional, Integral), Línea (Expresión, Gestión, Historia, Integral, Metodología, Taller

de Diseño, Tecnología, Teoría), área de concentración (Artes Gráficas, Disciplinares, Diseño Tridimensional, Divisionales, Ilustración, Medios Audiovisuales, Medios Digitales, Movilidad, Prácticas Profesionales, Tipografía), objetivo general y parciales, perspectiva desde la sustentabilidad, contenido sintético, modalidades de conducción, modalidades de evaluación y bibliografía. Se agregan siete campos auxiliares con fines de localización y ordenamiento, como nombres de archivo y longitudes de los textos. La extensión de un programa era entre una y cuatro páginas.

En este diagnóstico se disponen los PE en forma tabular de 21 columnas y 99 renglones convirtiéndolos en una base de datos en Excel; cada fila se llama una instancia, cada columna es uno de los componentes listados antes. Esta matriz puede ordenarse o filtrarse por diversos criterios.

Análisis general

De ese catálogo se seleccionan las columnas UEA, objetivo general, objetivos parciales, perspectiva sustentabilidad, contenido sintético, modalidades conducción, modalidades evaluación, se trasladan a Microsoft Word, acumulándose 28 373 palabras; se eliminan preposiciones y locuciones prepositivas, artículos, números, símbolos u otras voces irrelevantes, quedando 16 033 palabras; se aplica corrección ortográfica al conjunto y para fusionar duplicados los plurales se convierten a singulares y los femeninos a masculinos³; el bloque resultante se traslada al programa Primitive Word Counter, que reporta 2 178 palabras únicas, siendo las treinta más frecuentes las siguientes:

Palabra	Cant	Palabra	Cant	Palabra	Cant
1 diseño	377	11 exposición	148	21 forma	108
2 evaluación	293	12 reporte	144	22 sistema	105
3 investigación	265	13 práctico	134	23 examen	104
4 ejercicio	225	14 entrega	126	24 recuperación	94
5 gráfico	197	15 alumno	121	25 desarrollo	93
6 proyecto	188	16 presentación	120	26 comunicación	92
7 trabajo	164	17 análisis	118	27 digital	91
8 global	162	18 profesor	110	28 proceso	86
9 clase	151	19 medio	109	29 visita	81
10 final	149	20 realización	109	30 parte	80

Tabla 2. Las 30 locuciones más frecuentes en el conjunto de los PE. Fuente: elaboración del autor.

Con el orden obtenido de los términos, una interpretación de esta enumeración sería “en DCG, se enfatiza la evaluación y la investigación, con ejercicios, proyectos y trabajos globales o finales y en clase, y la entrega y presentación por el alumno de reportes con su práctica y análisis, tras la exposición del profesor acerca de medios, realizaciones, formas, sistemas; podrá haber exámenes de recuperación, ocasionalmente se desarrollarán procesos digitales y visitas”.

³ Así, pesa lo mismo “básico” que básica, básicas, básicos.

Esto apunta a una orientación pedagógica básicamente tradicional, tendiente a la exposición presencial de contenidos por el profesor; la investigación y análisis del alumno, y su evaluación global o final mediante ejercicios, examen, trabajos, proyectos y reportes en clase o entrega. Nótese que el término *digital* aparece hasta el puesto 27.

Ánalisis de objetivos

El cuerpo de cada PE incluye un objetivo general de entre diez y ochenta⁴ voces, entre uno y siete objetivos parciales⁵ en forma de balas o lista numerada de longitud variable. Se repite el procedimiento previo utilizando exclusivamente esas columnas. En conjunto quedan 1 343 palabras únicas, las treinta más frecuentes son:

	Palabra	Cant		Palabra	Cant		Palabra	Cant
1	diseño	138	11	medio	33	21	principio	23
2	gráfico	95	12	analizar	32	22	producción	23
3	comunicación	52	13	básico	31	23	desarrollo	22
4	aplicar	49	14	proyecto	30	24	solución	22
5	identificar	46	15	elemento	29	25	conocimiento	21
6	conocer	40	16	creación	28	26	problema	20
7	proceso	39	17	imagen	26	27	aspecto	20
8	sistema	38	18	lenguaje	26	28	específico	20
9	técnico	38	19	forma	24	29	producto	20
10	digital	36	20	objeto	23	30	utilizar	20

Tabla 3. Los 30 vocablos más frecuentes en los objetivos. Se resaltan los verbos. Fuente: preparación del autor.

En la serie anterior aparecieron sólo cinco verbos. Extendiendo la lista a los primeros nueve verbos en infinitivo (de un total de 123) que aparecen en el conjunto de objetivos se acumulan 248 apariciones que representan el 44% del total de 561 ocurrencias; se muestran en la tabla 4.

	Verbo	Cant		Verbo	Cant		Verbo	Cant
1	aplicar	49	4	analizar	32	7	desarrollar	16
2	identificar	46	5	utilizar	20	8	distinguir	16
3	conocer	40	6	comprender	17	9	explicar	12

Tabla 4. Primeros nueve verbos en las columnas de objetivos. Fuente: hechura propia.

Visto como catálogo, el apareamiento de los principales verbos de los PE, cuyos tamaños se muestran proporcionales al número de veces que aparecen, contra la Taxonomía de Bloom queda como se ve en la enumeración siguiente:

⁴ Ética [1402062] y Razonamiento y Lógica Simbólica I [1400042] respectivamente.

⁵ Comunicación y Diseño III [1420057] y Cultura y Diseño II [1400030] respectivamente.

Orden alto	6	Crear	desarrollar		
	5	Evaluar	explicar		
	4	Analizar	anализar		
	3	Aplicar	aplicar	utilizar	
	2	Entender	identificar	comprender	distinguir
	1	Recordar	conocer		

Tabla 5. Apareamiento de verbos en los objetivos de los PE en la Taxonomía de Bloom. Fuente: preparación de Torres, F.

La conclusión que se obtiene es que el grueso de los verbos de los objetivos de los PE cae en la parte media-baja de los niveles cognitivos de la taxonomía.

Análisis de contenidos sintéticos

El siguiente campo en los PE son los contenidos sintéticos. Con 20 voces, la UEA con dicho campo más breve es Expresión del Diseño Gráfico IV (Fotografía básica) [1402059], y el más extenso (450 voces) es Azoteas verdes [1401064]. Se aplica un procedimiento similar; se espera cierta dispersión por los múltiples temas que cubren las UEA. Esto se observa en la tabla 6:

	Palabra	Cant		Palabra	Cant		Palabra	Cant
1	diseño	70	11	análisis	23	21	tipografía	18
2	color	40	12	concepto	23	22	tipo	17
3	gráfico	38	13	investigación	23	23	elemento	16
4	medio	34	14	comunicación	21	24	espacio	16
5	imagen	30	15	desarrollo	21	25	uso	16
6	proceso	27	16	proyecto	21	26	identidad	15
7	digital	25	17	arte	20	27	información	15
8	característico	24	18	forma	20	28	relación	15
9	producción	24	19	básico	19	29	técnico	15
10	sistema	24	20	visual	19	30	discurso	14

Tabla 6. Los 30 términos más frecuentes en contenidos sintéticos. Fuente: elaboración de Torres, F.

Análisis de conducción del proceso

El siguiente campo es conducción del proceso de enseñanza-aprendizaje. El procedimiento arroja esta tabla:

	Palabra	Cant		Palabra	Cant		Palabra	Cant
1	exposición	88	11	documental	44	21	proyecto	26
2	profesor	80	12	visita	38	22	distancia	25
3	alumno	72	13	práctico	37	23	diseño	25
4	parte	58	14	curso	32	24	realización	24
5	investigación	58	15	trabajo	31	25	audiovisual	23
6	temático	57	16	interés	30	26	demostración	22
7	análisis	53	17	virtual	29	27	campo	22
8	presentación	49	18	presencial	29	28	elaboración	21
9	asesoría	46	19	sitio	27	29	conducción	21
10	ejercicio	45	20	semipresencial	26	30	apoyo	21

Tabla 7. Las 30 voces más frecuentes en el apartado conducción. Fuente: confección del autor.

Se colige que la modalidad más utilizada de conducción del proceso es la exposición del profesor al alumno, y las actividades más comunes son la investigación documental, el análisis temático, la asesoría y la presentación de ejercicios, las visitas a sitios de interés y el trabajo práctico en el curso de modo presencial, con alguna actividad virtual o semipresencial.

Análisis de evaluación del proceso

El siguiente campo es evaluación del proceso, el resultado es:

	Palabra	Cant		Palabra	Cant		Palabra	Cant
1	evaluación	270	11	recuperación	94	21	tiempo	46
2	global	160	12	trabajo	92	22	forma	45
3	investigación	156	13	práctico	81	23	presentación	44
4	reporte	142	14	complementario	74	24	ensayo	42
5	ejercicio	133	15	participación	67	25	parcial	39
6	entrega	123	16	realizado	64	26	exposición	37
7	final	123	17	realización	60	27	periódico	37
8	clase	122	18	inscripción	54	28	visita	36
9	examen	104	19	previo	54	29	extra	34
10	proyecto	97	20	requiere	54	30	grupal	34

Tabla 8. Los 30 términos más frecuentes en proceso de evaluación. Fuente: hechura de Torres, F.

Análisis de bibliografía

El campo que sigue es la bibliografía, de entre tres y 41 libros⁶, que en el formato APA se subdivide en autores, fecha de publicación, títulos, ciudades, editoriales y ligas de internet. Los contenidos originales son desiguales, no todos siguen las normas APA y tienen errores de ortografía y orden. Se limpiaron, ordenaron y colocaron en formato tabular, para proceder al cotejo; de los 99

⁶ Fotografía Básica [1402059] y Cultura y Diseño II [1400030] respectivamente.

PE se obtienen 1 021 entradas bibliográficas, los apellidos de los autores son muy variados, siendo el más común *Martínez*.

Por fecha: solamente 965 de las 1 021 referencias están fechadas (94.5%). Los cinco años más citados son 2002, 2009, 2003, 2008, 2007; la más antigua es de 1934 y la más reciente 2015.

Dado que los PE son de 2016, el total se agrupa en cuatro rangos: entre 20 y 80 años de antigüedad (1936-1996), entre 10 y 19 (1997-2006), entre cinco y nueve (2007-2011) y hasta cuatro años (2012-2015).

Antigüedad	Años	Referencias	Porcentaje	Acumulado
20 a 80	1936-1996	309	32.0%	32.0%
10 a 19	1997-2006	398	41.2%	73.2%
5 a 9	2007-2011	225	23.3%	96.5%
hasta 4	2012-2015	33	3.4%	100.0%

Tabla 9. Antigüedad de la Bibliografía en los 99 PE. Fuente: confección del autor.

Aquí se ve que 32% de los libros tienen más de veinte años y casi tres cuartas partes una década o más de haber sido publicados. Solamente el 3.4% son libros con menos de cinco años.

Por títulos: también son variados; los 5 términos más frecuentes son: diseño, design, arte, manual e historia; las cinco frases más frecuentes son diseño gráfico, geometría descriptiva, guía completa, artes gráficas y graphic design. Esto se aprecia en la [ilustración 1](#).

Por ubicaciones: las cinco más citadas son México, Barcelona, Madrid, España y Nueva York.

Por editorial: por mucho la más mencionada es Gustavo Gili, luego Hermann Blume, McGraw Hill, Paidós, UAMA, UAM, UNAM y Alianza.

Por direcciones de Internet: es la categoría más exigua. En 1 021 entradas solamente aparecen ligas electrónicas 39 veces (5 de ellas duplicadas). Además no están distribuidas, se concentran tan sólo en trece⁷ UEA (13.1% de los 99 PE). Lo más frecuentes es: Adobe, Encuadre, Illustrator, Inkscape, WordPress y Gimp, apuntando a sitios genéricos como adobe.com, www.aiga.org o www.signs.org; asoma un solo archivo PDF⁸, y ni un solo artículo o revista indexada.

Análisis de perspectiva desde la sustentabilidad

Sólo 15 de los 99 PE lo contienen, con sólo dos redacciones casi idénticas. En resumen dice:

⁷ Dichas UEA son: Tecnología para el Diseño Gráfico I (Software Básico), Tipografía Digital, Diseño de Mensajes Gráficos IV (Sistemas de Signos, Publicaciones), Diseño de Mensajes Gráficos VII (Sistemas de Signos, Orientación Espacios), Tecnología para el Diseño Gráfico V (Páginas Web), Diseño del Envase y Embalaje, Oralidad y Escritura para Diseñadores, Teoría y Metodología Aplicada I, Ilustración Expresiva, Infografía, Multimedia, Manipulación Digital, Diseño de Interfaz.

⁸ Dicho PDF es *Milestones in the history of thematic cartography, statistical graphics, and data visualization* (2009) <http://www.math.yorku.ca/SCS/Gallery/milestone/milestone.pdf>

Aplicar medidas y materiales sustentables o por lo menos cuidadosos con el medio ambiente en la realización de los ejercicios y en el desarrollo de la profesión: papel reciclado para bocetos y ejercicios de entrenamiento, utilización responsable del agua y la energía, limpieza de instrumentos, mesas de trabajo, restiradores y pisos, con detergente, fibras y trapos, no permitir el uso de papel higiénico, separación de residuos dentro de los tambos destinados a su recolección. Fomentar estas acciones aún fuera de la universidad (CoDo, 2016).

Conclusiones y propuestas

Los PE adolecen de algunas deficiencias que pueden solventarse, por ejemplo:

Fijar extensiones mínima y máxima para todos los apartados, que den homogeneidad y claridad a cada propuesta.

Redactar los objetivos en un formato con metas más claras, observables y medibles, utilizando un enfoque por competencias que incluya competencias blandas, apoyándose en la Taxonomía de Bloom para la era digital y procurando incluir niveles cognitivos más elevados.

En los modos de conducción, incluir explícitamente procesos con TIC que potencien el desarrollo, más allá de las herramientas propias del diseño, por ejemplo uso de blogs y foros, videoconferencias, podcasts, sistemas de mensajería, dispositivos móviles, etc. incrementando modalidades alternativas, como al aula invertida.

En la evaluación, privilegiar el uso de rúbricas, que hagan explícito lo que se espera del alumno y los valores parciales y totales de cada entregable.

En la bibliografía, fijar números mínimo y máximo de referencias, jubilar fuentes antiguas remplazándolas con otras más recientes y acordes con la realidad actual, especialmente documentos electrónicos apropiados, en todos los casos ajustando su citación al formato APA.

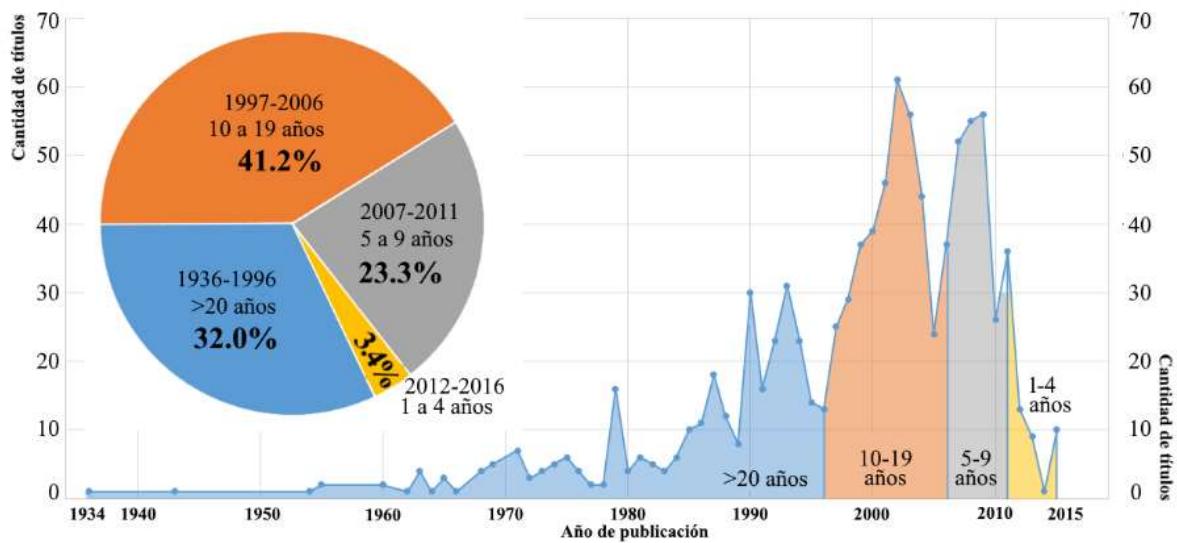
Efectuar una revisión ortográfica y de estilo con el fin de uniformar los PE antes de publicarlos, y luego hacerlos disponibles a la comunidad en un formato electrónico accesible.

Acometer los pasos pertinentes para que los PE puedan ser revisados y actualizados dinámicamente al menos cada año.

Referencias

- CoDo. (octubre de 2016). *Paquete propuesta de planes y programas de estudio CyAD UAM-Azc*. Recuperado el 29 de octubre de 2016, de Coordinación de Docencia: <http://bit.ly/2eA6xxV>
- Heer, R. (enero de 2012). *Revised Bloom's Taxonomy*. Recuperado el 20 de noviembre de 2017, de Iowa State University Center for Excellence in Learning and Teaching: www.celt.iastate.edu/teaching/RevisedBlooms1.html
- Microsoft. (14 de marzo de 2017). *Conceptos de minería de datos*. Recuperado el 02 de 01 de 2018, de SQL Server Analysis Services: <https://docs.microsoft.com/es-es/sql/analysis-services/data-mining/data-mining-concepts>
- Miller, S. (10 de febrero de 2018). *PureText 6.1*. Recuperado el 06 de marzo de 2018, de stevemiller.net/PureText/
- Montes-y-Gómez, M. (abril de 2003). *Minería de texto: Un nuevo reto computacional*. Recuperado el 06 de diciembre de 2017, de Instituto Nacional de Astrofísica, Óptica y Electrónica: <https://ccc.inaoep.mx/~mmontesg/publicaciones/2001/MineriaTexto-md01.pdf>
- Primitivezone. (2013). *Primitive Word Counter - Simple tool for counting keyword density in a text*. Recuperado el 29 de octubre de 2016, de Primitivezone.com: <http://www.primitivezone.com/primitive-word-counter.html>
- Zygomatic. (julio de 2015). *WordClouds*. Recuperado el 02 de febrero de 2018, de wordclouds.com: wordclouds.com

Ilustraciones



Francisco Torres García, estudiante del Doctorado en Diseño y Visualización de la Información en CyAD de la UAM-A; Maestro en Comunicación y Tecnologías Educativas por el ILCE; Ingeniero Mecánico Electricista (UNAM); profesional certificado Microsoft, evaluador del Sistema Nacional de Competencias.

Anteriormente ha sido: becario de la Fundación TelMex; académico de la UNAM en Matemáticas Aplicadas, Actuaría e Ingeniería; miembro de Asociación Latinoamericana de Profesionales en Seguridad Informática; presidente de la Asociación Mexicana para el Fomento del Libro Infantil y Juvenil. Promueve la cultura informática y el cómputo educativo; brinda capacitación de usuarios e instalación y mantenimiento de computadoras y redes.

