

4 La disciplina de la codificación del conocimiento.

4.1 Concepto de la codificación del conocimiento

Podemos considerar que "codificar" significa transformar el conocimiento en un código que permita organizarlo, hacerlo explícito, transportable y fácil de entender.

La codificación del conocimiento hace referencia a la habilidad de la empresa para estructurar el conocimiento en un conjunto de reglas y relaciones identificables que puedan ser fácilmente comunicables, por lo que la codificación de conocimiento puede entenderse como la representación del conocimiento de tal manera que éste pueda ser accedido y transferido.

David y Foray (1995) definen la codificación del conocimiento como la transformación del conocimiento en información que puede ser fácilmente transmitida a través de las infraestructuras de la información. Es un proceso de reducción y conversión que hace especialmente fácil la transferencia, verificación, almacenamiento y reproducción del conocimiento.

La codificación se fundamenta en almacenar conocimiento explícito en bases de datos o sistemas expertos, de manera que éste pueda ser utilizado con posterioridad.

Conwan and Foray (1997) postulan que la codificación del conocimiento es el proceso de conversión del conocimiento en mensajes que pueden ser procesados como información.

La codificación es un proceso que implica la creación de modelos, hay un proceso de creación, no es solo una traducción.

4.1.1 Los códigos

Cowan, David, y Foray (2000) recalcan que el significado del conocimiento codificado no coincide totalmente con el de conocimiento explícito. Lo más obvio es relacionarlo con la presencia de algún código, entendido como un conjunto de notaciones, símbolos, reglas, un género de diccionario aceptado como autoridad por su uso y por el consentimiento general. La actividad inicial de codificación siempre implica crear primero un diccionario especializado. Es interesante recordar que Milton Friedman (1935) señalaba que una parte de la economía positiva podía ser vista como un lenguaje y que su función es servir de sistema de clasificación para organizar material empírico y facilitar nuestro entendimiento sobre él. Codificar una porción de conocimiento significaría añadir contenido a un código, que es elaborado sobre la base de otro contenido previamente existente. El conocimiento registrado en un código constituye un género de depósito, punto de referencia o de autoridad. Pero la posibilidad de poder acceder exitosamente al mismo implica la necesidad de haber realizado una adquisición previa de conocimiento especializado, en algunos casos de considerable magnitud y profundidad. Sólo un grupo determinado que haya alcanzado ciertos niveles de

conocimientos podrá acceder a los códigos especializados. Para ellos el conocimiento será codificado, para otros se tratará de conocimiento tácito y, para la mayoría, en caso de conocimientos muy avanzados, se tratará de cuestiones totalmente incomprensibles.

Por tanto, la codificación de la información necesita la aparición y la definición de los códigos, lo que tienen una gran importancia en la transferencia del conocimiento a individuos y grupos, pues han de saber cómo interpretar y decodificar la información.

Tanto la información como el conocimiento deben ser claros y entendibles para todos los usuarios reales y potenciales. En algunas ocasiones cuando se juntan un grupo de expertos en áreas precisas como medicina, marketing, producción, petróleo, química, finanzas, legales, sistemas u otros, manejan códigos que ellos solos entienden.

La tarea de los administradores o gestores del conocimiento es "bajar" el conocimiento a un lenguaje o código que sea familiar y comprensible para el resto de los interesados.

La elección del código depende de los destinatarios y usuarios que tiene como objetivo el conocimiento. Con esto no alcanza solamente con codificarlo, sino que hace falta tener en cuenta quienes van a ser los destinatarios de dicho conocimiento.

Un mensaje es expresado en un cierto código (algún lenguaje) y en principio puede ser almacenado, al menos temporalmente, en algún medio (papel, dvd, cd ...) esto no elimina la relación entre: medio, mensaje y lenguaje.

4.1.2 Principios de la codificación del conocimiento

La codificación en las organizaciones adapta el conocimiento en formatos accesibles y aplicables. Por lo tanto las empresas que deseen codificar el conocimiento exitosamente deben tener en cuenta estos cuatro principios:

- Los administradores deben decidir a qué objetivos de la organización servirá el conocimiento codificado.
- Para alcanzar estos objetivos, deben ser capaces de identificar el conocimiento existente en distintas formas.
- Los administradores del conocimiento deben evaluar la idoneidad y utilidad del mismo para su codificación.
- Los codificadores deben identificar un medio adecuado para la codificación y distribución.

4.1.3 Procesos de codificación

En la literatura aparecen dos aspectos del proceso de creación de la información:

- La codificación del conocimiento tácito previo.

Cook y Brown (1999) postulan la existencia de este tipo de codificación, definiéndola como el proceso mediante el cual el conocimiento se hace explícito. Es lo que Nonaka and Takeuchi (1995) hablan en su distinción de un movimiento entre formas tácitas y explícitas de conocimiento. La codificación del conocimiento en este sentido debe conllevar el uso del lenguaje para articular, describir, explicar... Quizás pueda ser visto como un nivel primario de codificación, no necesariamente realizado al principio, pero primario en el sentido que la codificación emerge de la persona que tiene el conocimiento.

- Codificación de información.

Este caso nos referimos al conocimiento que ya ha sido codificado en información. En este contexto, Sorensen y Lundh-Snis (2001) identifican la codificación de la información como el proceso de clasificación de la información, mientras que Baumard (1999) habla sobre la codificación de los datos de investigación. Ésta se puede considerar como una codificación secundaria, donde la codificación es del conocimiento que ya ha sido codificado como información (podemos todavía llamar a esta información “conocimiento” codificado desde la perspectiva de la persona responsable de su codificación primaria, aunque es “información” desde la perspectiva del resto de la organización).

4.2 Objetivos de la codificación del conocimiento

El objetivo de la codificación consiste para Davenport y Prusac (2001) en poner el conocimiento institucional en alguna forma legible, entendible y organizada, al alcance de quienes lo necesiten. Literalmente convertir el conocimiento en un código (aunque no necesariamente en un código informático) para que sea lo más organizado, explícito, portátil y fácil de entender posible.

Se trata de hacer visible y accesible el conocimiento institucional, con el propósito de que sea usado para realizar las decisiones.

La codificación del conocimiento es una actividad clave en la construcción de la base mínima de conocimiento, en el fortalecimiento del proceso de creación de conocimiento y en la construcción de capacidades tecnológicas centrales.

4.2.1 El reto de la codificación

Un reto de la codificación radica en que es imposible hablar sobre conocimiento en términos de unidades discretas. Por estas razones, revisar y categorizar conocimiento es una tarea difícil.

La codificación de todo el conocimiento de la organización sería una tarea inmensa e inútil, similar e incluso más difícil que los esfuerzos en general adoptados para desplegar el modelo de datos de toda la organización. Debido a que el objetivo de la codificación consiste en poner el conocimiento en una forma en que se le pueda utilizar, la organización necesita tener cierta idea de cuales son los usos que le serían destinados.

Sin embargo para que valga la pena un proyecto de codificación del conocimiento necesita objetivos más específicos que simplemente hacer que el conocimiento esté disponible en general. Davenport (2001) razona que es esencial encontrar las fuentes del conocimiento que uno desea codificar. Si uno no sabe dónde se encuentra, no puede hacer nada con éste y es imposible saber de qué se trata.

Uno de los retos más grandes en la administración del conocimiento es documentar todo ese conocimiento que tienen los empleados en las organizaciones, tal y como lo dice Davenport (2001): “el conocimiento tácito, complejo, desarrollado y asimilado por las personas durante largo tiempo es prácticamente imposible de ser reproducido en un documento o base de datos. Dicho conocimiento incorpora tanto aprendizaje acumulado y arraigado que puede ser imposible separar las reglas en que actúa un individuo. De igual manera, el conocimiento utilizado por un científico investigador creativo, para decidir la línea de investigación a seguir, no se puede convertir en una lista o informe detallado; si fuera posible extraer de esta manera el conocimiento del individuo, cambiarían radicalmente las políticas de compensación y educación”.

Por este motivo, el proceso de codificación del conocimiento tácito más vasto en las organizaciones generalmente se ve limitado a encontrar a una persona que tenga el conocimiento, dirigir a quien está buscando hacia la misma que tenga el conocimiento y alternarlos para que interactúen.

4.2.2 Importancia de la codificación

Independientemente de las intenciones que motivan el esfuerzo de codificación, el proceso por el cual estos medios se crean y se actualizan implica generalmente un esfuerzo para entender las relaciones causales entre las decisiones tomadas y los resultados esperados en el rendimiento, aunque el aprendizaje pudiera no ser el objetivo deliberado de la codificación.

Por otra parte, Schulz y Jobe (2001) apuntan que la transmisión eficiente del conocimiento tácito requiere su codificación en formas explícitas. Así, la codificación facilita los flujos de conocimiento organizacional, y estos flujos son importantes porque transmiten el “know-how” dentro de la organización, facilitan la coordinación y permiten el reconocimiento y la explotación de economías de escala. Además, el carácter global del mercado demanda un incremento del intercambio de conocimiento

entre subunidades geográficamente dispersas y, por tanto, incentiva a las empresas a invertir en procesos de codificación.

La codificación de conocimiento es un elemento importante y relativamente poco reseñado en la creación de capacidades dinámicas. La literatura ha subrayado que la codificación facilita la difusión del conocimiento existente (como por ejemplo Nonaka en 1994), así como la coordinación y realización de actividades complejas. Habiendo identificado y seleccionado el cambio de las rutinas operativas o el establecimiento de nuevas rutinas, la organización debería crear un manual o un instrumento para facilitar su réplica y su difusión. Las ventajas principales del esfuerzo de codificación surgen del uso acertado del manual o instrumento. Para desarrollar un manual para la ejecución de una tarea compleja, los individuos que se implican en el proceso deberían formar un modelo mental sobre qué acciones deben ser ejecutadas y bajo qué condiciones. De la realización de dicho esfuerzo, es probable que se evolucione con una definición más precisa sobre qué funciona, qué no funciona y por qué.

La codificación es, por tanto, potencialmente importante como mecanismo de apoyo en el proceso completo de la evolución del conocimiento, y no constituye solamente una fase de transferencia. La codificación puede facilitar, por ejemplo, la generación de nuevas propuestas para cambiar las rutinas actualmente disponibles, así como la identificación de fuerzas y debilidades de nuevas propuestas al actual conjunto de rutinas. La simplificación cognoscitiva inherente en el acto de sintetizar sobre el papel (o en un programa informático) la lógica que hay tras un conjunto de instrucciones puede representar:

- por un lado, tal y como remarca Boicot (1998) un ahorro en proceso de datos que permita una toma de decisiones más eficaz
- Weick, (1995) apunta, por otro lado, que también puede representar una forma más o menos deliberada de creación de significado (“sensemaking”) retrospectivo sobre el rendimiento obtenido de un conjunto de actividades. Esta creación de significado se suele representar mediante un lenguaje especializado y un entendimiento causal desarrollado para y por otros esfuerzos de codificación de un tipo similar.

4.2.3 Beneficios de la codificación

A parte de los beneficios ya comentados en apartados anteriores cabe destacar expresamente los siguientes:

- Entrenamiento e instrucción. Se proporciona entrenamiento del personal más novato basándose en la codificación del conocimiento de los empleados de mayor rango o antigüedad.
- Predicción. Se infieren los resultados más probables dada una situación, y se muestra rápidamente los avisos adecuados o las sugerencias de medidas correctoras.

- Diagnósticos. Se localizan los síntomas identificables de los factores causales específicos.
- Planificación y programación. Se organiza el transcurso total de una acción antes de que se tome ningún paso.

4.3 La codificación de los diferentes tipos de conocimiento

Los distintos tipos de conocimiento se prestan de manera diferente a su codificación. Distintos tipos de conocimiento implican distintos tipos de lenguaje (música, matemáticas...). La comunidad a la que va dirigida ha de conocer el vocabulario y los conceptos del lenguaje.

En este sentido Davenport y Prusak (2001) reconocen siete dimensiones en el eje de la codificación, a saber:

- imposible / posible de ser enseñado
- no articulado / articulado
- no observable / observable
- rico / esquemático
- complejo / simple
- no documentado / documentado
- tácito / articulado

Uno de los aspectos importantes de la codificación que señalan Schulz y Jobe (2001) es que cada tipo de conocimiento requiere una forma distinta de codificación, de tal forma que una adecuada codificación facilitará el flujo intraorganizativo de conocimiento. Así, el conocimiento técnico puede ser codificado en números y códigos, el conocimiento de marketing y ventas puede ser codificado en texto y formas basadas en el lenguaje, y el conocimiento estratégico respecto a los competidores, es depositado en objetos y personas. El conocimiento codificado en tales combinaciones facilita los flujos de conocimiento y repercute positivamente en los resultados de las empresas.

En estudios realizados en la OECD (1996) se distinguen cuatro tipos de conocimiento:

- **Know-what.** El “conocer que” (know-what) es el tipo de conocimiento proposicional que se refiere a la información o conocimiento acerca de hechos. Es el concepto de conocimiento que más se acerca al de “información” y es el más fácil de ser transferido.

- **Know-why.** El “conocer porque” (know-why) es el conocimiento de cómo las cosas funcionan, el conocimiento de los principios y leyes que rigen tal funcionamiento. Este tipo de conocimiento ha sido muy importante para el desarrollo tecnológico.
- **Know-how.** El “conocer como” (know-how) se refiere a los talentos, habilidades y competencias de las personas. Puede tratarse en muchos casos de conocimiento tácito, difícil de codificar y transferir. Con frecuencia este tipo de conocimiento no puede ser separado de la persona u organización que lo posee. Este sería el conocimiento definido como “sticky” por von Hippel (1994).
- **Know-who.** El “conocer quien” (know-who) se refiere a relaciones de conocimiento acerca de quién conoce qué y quién conoce como. En un estudio, realizado por Foray y Lundvall (1996) para la OECD, sus autores identifican los dos primeros tipos con el conocimiento codificado y los dos últimos con el conocimiento tácito.

Lundvall y Jonson (1994) analizan la codificación de los cuatro tipos de conocimiento que identifican; que a nivel de individuo son el know-what, know-why, know-how y el know-who, que a nivel de compañía se corresponde con bases de datos, modelos de interpretación, rutinas y redes respectivamente.

4.3.1 La codificación del know-what

Se define como el conocimiento de los hechos. Aquí el conocimiento está cercano a lo que normalmente se denomina información, se puede subdividir y comunicarse como datos. Es relativamente fácil de codificar. Las bases de datos dan soporte a este tipo de conocimiento.

La tecnología de la información pone al servicio de los individuos esta información, aunque esta información tiene que ser encontrada y lo que es más relevante, ha de ser seleccionada y usada. La efectividad de los motores de búsqueda especifican cómo de accesible es la información. Aún con los últimos avances el acceso a esta información es imperfecta y en muchas ocasiones hemos de acceder por medio del Know Who. Incluso con internet la información es fácil de encontrar y acceder, pero su ruta de acceso puede desaparecer en un futuro.

4.3.2 La codificación del know-why

Conocimiento de los principios y leyes en la naturaleza, sociedad y mentalidad humana. Es extremadamente importante en el desarrollo tecnológico en la ciencia base (química, electrónica...). El acceso a este tipo de conocimiento acelerará el avance tecnológico y reducirá la frecuencia de errores. La codificación aquí es normalmente incompleta (la ciencia base se construye parcialmente sobre las habilidades personales).

El trabajo científico se centra en producir modelos teóricos para este tipo de conocimiento, algunos de estos trabajos se encuentran en sitios de dominio público. Los

académicos tienen fuertes incentivos para publicar y hacer accesibles sus resultados. Internet ofrece una gran velocidad de propagación por medio de la publicación electrónica, esto no quita que en ocasiones haya que invertir esfuerzo de aprendizaje antes de que la información tenga algún significado.

Para tener acceso al know-why científico es necesario perseguir actividades de I+D e invertir en ciencia. Esto es cierto para individuos, regiones y empresas. En actividades industriales tecnológicamente muy competitivas las soluciones técnicas van por delante del know-why académico, en estos casos se pueden resolver problemas sin un claro conocimiento de “cómo funciona”. Aquí el conocimiento es más know-how que know-why.

4.3.3 La codificación del know-how

Se refiere a las habilidades de la persona, “saber hacer”. Parte de este conocimiento puede ser posible de articular en partes para ser codificado, pero siempre nos vamos a encontrar diferencias entre la habilidad de un cirujano y el libro codificado que usa. Incluso el uso de las operaciones matemáticas más elementales pueden depender de la experiencia de la persona.

A medida que la complejidad del conocimiento base aumenta, la cooperación entre organizaciones se hace más interesante y necesaria.

Una de las razones más importantes para las redes industriales es la necesidad para las firmas de ser capaces de compartir y combinar elementos de Know-How. Es el conocimiento donde la tecnología de la información se encuentra con mayores problemas a la hora de transformar el conocimiento tácito en explícito codificado. Un cocinero puede escribir un libro de recetas, pero lo que hagamos nosotros con ella puede no parecerse al resultado deseado.

Los intentos de la tecnología de la información en desarrollar sistemas expertos da muestra de lo difícil y caro que puede resultar el transformar habilidades en información que pueda ser usada por otras personas. Hatchuel y Weil (1995) demostraron que desarrollar un sistema experto implica cambios en el contexto del conocimiento del “experto”. Esto no es sólo cierto para habilidades personales, sino también para las profesionales y de equipo.

Las redes electrónicas no pueden sustituir las relaciones personales, los sistemas multimedia y realidad virtual combinadas con técnicas de telecomunicación pueden ayudar en la transferencia de este conocimiento.

4.3.4 La codificación del know-who

Implica conocimiento acerca de quién conoce el qué y quién conoce qué hacer, además de la capacidad de cooperar y comunicarse con diferentes tipos de personas y expertos. Es muy difícil de codificar.

Los expertos pueden ayudar a los amateurs a hacer comprensible el conocimiento. Esta puede ser una gran motivación para la participación de las universidades en las empresas, así mismo es un foco de participación de las empresas privadas en la investigación básica.

En muchos casos la explotación de nuevas ideas (y personas) por parte de las empresas minan el intercambio abierto que ha caracterizado a la producción de conocimiento de las universidades, (como sucede en biotecnología).

4.4 La codificación del conocimiento tácito.

El conocimiento también puede codificarse de manera tan articulada y clara que se puede expresar en un lenguaje específico y registrar en un medio particular. La codificación implica la exteriorización de la memoria. Depende de una gama de acciones que se van complicando, como usar un lenguaje coloquial para escribir una receta de cocina, aplicar técnicas de diseño industrial para dibujar un boceto a escala de una pieza de maquinaria, crear un sistema especializado con base en las reglas formales de la inferencia dando especial importancia a la serie de etapas orientadas al problema, entre otras. Como tal, el conocimiento es independiente del individuo y de la memoria, así como la capacidad de comunicación del ser humano (siempre que el medio en el que se almacena el conocimiento esté resguardado y el lenguaje en el que se exprese sea recordado).

4.4.1 Efectividad de la codificación

Con el surgimiento de la codificación, Goody (1977) razona que el problema de la memoria deja de dominar la vida intelectual. Así, se producen programas de aprendizaje que sustituyen de manera parcial a la persona que posee un conocimiento y que lo enseña. Goody (1977) pone el ejemplo de que una receta escrita puede llenar parcialmente el vacío producto de la ausencia de la abuela.

La palabra clave aquí es “parcialmente” porque la codificación equivale a un proceso en el que se reduce el conocimiento humano a información, y en el curso de tal transformación casi de hecho se alterarán algunas cosas de alguna manera y, muy probablemente, se perderán otros significados. Lo que se expresa y se registra, por tanto, no es conocimiento del todo. Es un programa de aprendizaje que ayuda a estabilizar y a reproducir el conocimiento. Cuando un joven técnico recibe un manual del usuario, no se le está dando directamente el conocimiento de “cómo usar el aparato”. Se confía en que el manual es útil y le servirá para reducir los costos de la reproducción de conocimiento. En muchos casos, cuando los técnicos han “aprendido a aprender” y están trabajando con una máquina más o menos estándar, la reproducción de conocimiento se vuelve casi instantánea y adquiere características cercanas a la repetición de información.

Sin embargo, en casos más complejos el conocimiento codificado, aunque ciertamente útil, sólo ayudará a medias. La reproducción de conocimiento se dará entonces por

medio de la capacitación, la práctica y las técnicas de simulación (pilotos aviadores, cirujanos).

4.4.2 Nuevas posibilidades cognitivas

Se debe destacar que hay una segunda función de la codificación que es crucial para los presentes propósitos. La codificación consiste en transcribir el conocimiento en representaciones simbólicas de tal manera que puedan almacenarse en un medio específico. Esto genera nuevas potencialidades cognitivas que seguirán siendo incalculables ya que el conocimiento está vinculado a individuos como tales y, por tanto, sólo será escuchado (si se expresa verbalmente) o visto (si se pone en práctica) por medio de la interacción con quienes lo poseen.

Plasmar el conocimiento (por medio de la escritura, la gráfica, el modelado o la virtualidad) posibilita analizar y organizar el conocimiento de diferentes formas, así como aislar, clasificar y combinar distintos elementos. Esto lleva según Goody (1977) a la creación de nuevos objetos del conocimiento, como listas, tablas, fórmulas, etcétera, los cuales son de importancia fundamental, puesto que dan pie a nuevas posibilidades cognitivas (clasificación, taxonomía, interconexiones múltiples, simulación) que pueden ofrecer un marco favorable para la pronta producción de conocimientos nuevos.

Sin embargo, sólo son factibles cuando la gente considera el hecho de registrarlos y, por consiguiente, la representación simbólica de sus estados cognitivos. Los avances metodológicos para registrar información basados en medios tecnológicos son de crucial importancia, ya que permiten que las representaciones del conocimiento pasen de la etapa denominada “prealfabetizada” (gestos y palabras) a la “alfabetizada” (escritura y dibujo) y por último a las etapas “posalfabetizadas” (diseño de interacciones estructuradas).

De esta forma, la codificación tiene un papel esencial en la economía del conocimiento porque está al servicio de la memorización, la comunicación y el aprendizaje futuros y establece una base firme para la creación de nuevos objetos

4.4.3 Una clasificación más flexible de conocimiento tácito

La literatura acerca del conocimiento tácito y codificado brinda un conjunto de ideas interesantes sobre las cuales es necesario reflexionar, pero no proporciona un marco analítico adecuado para analizar empresas que presentan como un rasgo característico la falta de codificación de conocimiento, por lo tanto es necesario avanzar en su construcción.

Una característica importante en el caso de que una organización descuide la codificación de cuerpos substancias del conocimiento codificable, hace necesario usar una definición más flexible del conocimiento tácito que incluya dos categorías:

- Conocimiento tácito no codificable. Se trata de conocimiento estrictamente tácito.

- Conocimiento tácito codificable. En contraste con el anterior, es conocimiento tácito susceptible de codificación, pero aun no lo ha sido, se refiere al conocimiento verbalmente codificado y no documentado.

Asimismo, y es útil usar una definición más rígida del conocimiento codificado que incluya sólo los documentos escritos, el software, el equipo, las especificaciones, etc. Este marco analítico se describe en la siguiente figura:

| Conocimiento tácito | | Conocimiento codificado |
|---------------------|---|--|
| No codificable | Codificable Codificado verbalmente y no documentado | Documentado Documentos escritos, expresiones gramaticales, expresiones matemáticas, especificaciones, diseños, manuales, equipo, software |

Figura 3: clasificación más flexible de conocimiento tácito

La codificación verbal se considera como una codificación incompleta del conocimiento, porque es difícil recordar y explicar detalles en palabras, y también porque repetir las mismas ideas con el mismo nivel de precisión no es fácil.

4.4.4 La codificación en estrategias tecnológicas inestables

Cuando se está en un entorno inestable en cuanto a la estrategia tecnológica de la organización, se presentan una serie de problemas a la hora de tratar de codificar el conocimiento tácito. Dutrénit (2000) detectó dos problemas asociados a la estrategia tecnológica de una organización:

- la dualidad de la estrategia
- la inestabilidad del apoyo a cada una de estas

Si bien las dos suelen coexistir en paralelo, no suelen ser articuladas consistentemente.

La inestabilidad de la estrategia tecnología provoca dificultades para cambiar las rutinas de codificación del conocimiento e influye negativamente en la conversión del aprendizaje individual en organizacional.

Cada estrategia demanda un diferente tipo de conocimiento para ser compartido, codificado e integrado, y fomenta la existencia variada de rutinas. Por un lado, la estrategia de ser “tecnológicamente independiente” estimula una mayor preocupación por compartir y codificar el conocimiento para apoyar su proceso de creación. Por otro, la estrategia de seguidor rápido, que tiene un enfoque de corto plazo se suele mostrar menos preocupada por compartir y codificar, o promueve el compartir y codificar otro tipo de conocimiento. Así, la dualidad de la estrategia tecnológica alimenta la

coexistencia de una combinación de prácticas, mientras que la inestabilidad de la estrategia tecnológica hace más difícil cambiar las rutinas.

La inestabilidad de la estrategia tecnológica genera dificultades para consolidar las prácticas de compartir y codificar el conocimiento. También debilita los esfuerzos realizados para promover el aprendizaje organizacional, lo cual depende de establecer nuevas prácticas para compartir y codificar el conocimiento. La manera más fácil de tratar con la inestabilidad en la empresa es continuar con el antiguo modo de operar.

De hecho la actividad de codificación es vista como una etapa final pero secundaria del desarrollo de proyectos. Según el énfasis cambiante de la estrategia tecnológica, el personal suele ser movido a nuevas actividades antes de culminar la codificación de los proyectos desarrollados. De esta forma, la inestabilidad de la estrategia tecnológica hace más difícil que el personal internalizara los nuevos arreglos organizacionales establecidos, orientados a cambiar sus prácticas y fortalecer la codificación del conocimiento, y puede llegar a explicar porqué el personal técnico no llena las bitácoras de los proyectos a pesar de contar con ellas.

Resumiendo, la inestabilidad debilita los esfuerzos para compartir y codificar el conocimiento y así aprender al nivel de la organización, aumenta las dificultades que usualmente se confronta para cambiar las rutinas, de tal forma que influye negativamente en el proceso de implantación de las nuevas prácticas orientadas a incrementar la socialización y codificación del conocimiento.

4.5 Evolución de la codificación.

4.5.1 El progreso de la codificación en la organización

Adicionalmente, este marco analítico destaca el hecho de que hay cierto conocimiento no codificado en la actualidad que en principio podría serlo. El carácter tácito del conocimiento y su posibilidad de ser codificable cambia con el tiempo: lo que es no codificable este año puede llegar a serlo el próximo. A lo largo del tiempo, la cantidad de conocimiento se incrementa, se adquiere nuevo conocimiento tácito; parte del conocimiento tácito no codificable de hoy lo es en el futuro y la cantidad de conocimiento codificado aumenta. Sin embargo, como Cowan y Foray (1997) indicaron, la codificación nunca es completa, y algunas formas de conocimiento tácito siempre continuarán jugando un papel importante.

En empresas que descuidan la codificación, el nuevo conocimiento tácito codificable permanecerá seguramente sin codificar. “Cuánto” y “a qué tasa de velocidad” el conocimiento tácito codificable se codificará, depende de varios factores, tales como:

- Según Cowan y Foray (1997), de los costos y beneficios económicos de la codificación.
- Para Levitt y March (1996) depende de la medida en que se considere que la valorización del conocimiento puede llegar a ser pertinente en el futuro

- Leonard y Sensiper (1998) apuntan que dependerá de la ignorancia de la gente acerca de las dimensiones tácitas de su conocimiento.

A lo largo del tiempo es difícil saber qué conocimiento tácito permanecerá siempre estrictamente así, en cualquier periodo hay cierto conocimiento que es no codificable; el ejemplo común es cómo manejar una bicicleta, empero es difícil discriminar entre diferentes tipos de conocimiento tácito así como analizar la relación entre el tácito y el codificado. Esos aspectos requieren mayor análisis.

A partir del marco analítico propuesto este trabajo examina en detalle uno de los modos de conversión de conocimiento descritos por Nonaka y Takeuchi (1995), la conversión de conocimiento tácito en codificado (exteriorización o codificación).

4.5.2 Avances en la codificación

Conwan y Foray (1997) enumeran los avances que se producen en la disciplina de la codificación del conocimiento que ayudan a ir perfeccionarlo:

- Desarrollo de nuevos lenguajes (en los que se incluye la reconstrucción de lenguajes perdidos), que puede permitir la codificación del conocimiento previamente a los pensamientos inherentemente tácitos.
- Cambios en nuestra habilidad de crear modelos de fenómenos y actividades.
- Cambios en las tecnologías de codificar y decodificar, dentro de los lenguajes y modelos existentes.
- Innovación en las tecnologías de almacenamiento, grabación y difusión de los mensajes. Este punto no se considera que afecte a la codificación per se: los avances en la tecnología asociados con los medios (grabación y almacenamiento), así como los producidos en la tecnología de la distribución son complementarios al proceso de codificación. Sin embargo, son decisivos en el aumento del valor económico del conocimiento codificado para aquellos agentes que quieren codificar el conocimiento con el propósito de facilitar el almacenamiento, acceso o distribución.

A medida que el conocimiento a codificar se hace más complejo, para hacerlo útil la codificación demandará mejores sistemas de almacenamiento y de gestión (nuevas formas de codificación, sistemas de ayuda on-line, sistemas expertos...)

Los avances en la codificación son relevantes pero limitados a los tres primeros procesos enumerados anteriormente. En las últimas décadas se han descubierto muchos cambios tecnológicos que han tenido impactos significativos en muchas partes de la economía. Algunos de esos cambios tendrán un efecto directo en la economía de la codificación del conocimiento. Conwan y Foray (1997) identifica principalmente los siguientes:

- Las tecnologías de telecomunicación han cambiado drásticamente, con el desarrollo de la conmutación digital, que ha facilitado cambios en las clases de servicios ofrecidos y en los costes de ellos. Además, las tecnologías de transmisión, por ejemplo el desarrollo e implementación de la redes de fibra óptica, han aumentado ampliamente la capacidad de volumen de la infraestructura física, lo que reduce el coste de las transmisiones de voz y datos.
- Es un hecho común que el crecimiento en la velocidad de las unidades de procesamiento han cambiado las capacidades de los ordenadores. Esto ha hecho útiles los ordenadores en lugares donde nunca han estado antes. Y han introducido la posibilidad de usar nuevas técnicas para dirigir viejos problemas. La velocidad de procesamiento ha hecho posible solucionar, por ejemplo, funciones analíticas simbólicamente, lo que ha relajado las restricciones de los formularios funcionales que los científicos han encontrado previamente. Igualmente, ahora es posible, en los formularios aparentemente sin solución, hacer una sistemática y minuciosa exploración del bloqueo de una función o sistema para resolver el problema. Para los sistemas dinámicos, la simulación ha sido posible como una herramienta analítica. Ergas (1991) comenta que, conectada con esos desarrollos, y formando parte del éxito de nuevas técnicas de modelo y solución están los recientes desarrollos aplicados a las matemáticas teniendo que ver con las técnicas de solución y aproximación. Las técnicas de aproximación que hacen que el tiempo de la solución sea lineal en lugar de exponencial han expandido mucho más el ámbito para usar herramientas de cálculo como parte del conjunto de herramientas de modelos o análisis.
- Los avances en las tecnologías usadas para almacenar datos han sido significativos, ahora, por ejemplo, las unidades de disco tienen más capacidad de almacenamiento que las que tenían las unidades de cintas magnéticas hace dos décadas.

En el contexto de esta discusión, esos cambios han mejorado nuestra capacidad de modelar fenómenos más complejos. Los avances en el desarrollo de soluciones algorítmicas, aumenta la utilidad de las simulaciones y supone que podemos sistematizar mas información con tantas posibilidades que sus estructuras y relaciones con otras estructuras se aclaran. Esta capacidad de modelar mejor y modelar nuevas cosas supone que en principio podemos codificar más clases de conocimiento.

En general, sin embargo, codificar nuevas clases de fenómenos requiere nuevas herramientas de codificación, especialmente, nuevos lenguajes. Hasta cierto punto se han desarrollado modelos como parte de ejercicios y haciendo uso de estructuras de lenguaje existentes, pero hay un punto en el cual se deben desarrollar nuevos. Un ejemplo puede ser la reciente necesidad de desarrollar el lenguaje de ordenadores de inteligencia artificial. Una vez que esos nuevos lenguajes existen, como quiera que sea, los costes de aumentar más el ámbito sobre los cuales tiene lugar la tarea de codificación, son en si mismos menores. Esto implica que el coste de codificación desciende con esos desarrollos tecnológicos.

No obstante, esos cambios generalmente no son los suficientes para tener, en la práctica, un efecto espectacular en la codificación. Típicamente, cuando el conocimiento es

codificado se hace más complejo, o si el fenómeno es más complejo, para que la codificación sea útil, se requiere almacenamientos mejorados. Más fenómenos complejos implican que el conocimiento codificado contendrá, en general, más información, en el sentido del número de bits. En principio esto no es un problema, simplemente significa que se requiere mayor cantidad de medios de almacenamiento (mas papel o cintas magnéticas quizá). Pero las nuevas formas de codificación, sistemas expertos o sistemas de ayuda on-line por ejemplo, requieren un rápido acceso a una gran cantidad de variada información (el almacén del conocimiento codificado). Esto ha sido posible gracias al desarrollo de nuevos y mejores medios de almacenamiento, que ha dado como resultado una reducción de los costes de almacenamiento, una reducción en los costes de uso, y un incremento en la velocidad de acceso.

De esta manera observamos como el coste de codificación ha descendido y el valor del conocimiento codificado ha crecido a causa de ser menos caro de reproducir y más valioso (rápido por ejemplo) de usar.

No obstante, este enunciado referente al aumento del valor del material codificado debe ser interpretado cuidadosamente. El material impreso, si el papel en el que se imprime es relativamente bajo en contenido ácido, puede durar varios siglos. Físicamente se deteriora lentamente, y el lenguaje que contiene escrito se mantiene de tal manera que interpretarlo es relativamente sencillo. Lo mismo no es necesariamente cierto en los documentos almacenados en soportes magnéticos o incluso en los soportes ópticos. Los soportes magnéticos se deterioran relativamente rápido desde un punto de vista físico (aproximadamente de 5 a 10 años es la duración estimada), y con los cambios rápidos en las tecnologías de hardware y software, el lenguaje que se escribe en ellos puede desaparecer también rápidamente.

El problema proviene del hecho de que con las nuevas técnicas de información no almacenamos documentos, sino más bien una serie de instrucciones que deben ser interpretadas y cumplidas por una combinación pertinente de hardware y software antes de que la información que contienen sea usada por los humanos. Actualmente, todavía no ha sido puesta en práctica una fiable y relativamente normalizada forma de archivar la información originada de esa manera. Varias soluciones se han sugerido, pero según Cook (1995), ninguna está exenta de problemas. Esos problemas existen, pero no son insuperables. Ciertamente los costes de almacenamiento y recuperación de datos a corto plazo han disminuido. Si se encuentra una solución a las dificultades mencionadas, también disminuirán los costes de almacenamiento a largo plazo.

Una cuestión más es que las nuevas tecnologías de información han reducido los costes de difusión del conocimiento codificado. Uno de los grandes beneficios de la codificación es, por supuesto, la manera en la que se facilita la diseminación del conocimiento. Esto se ha producido claramente por el descenso en los costes de las telecomunicaciones. Esto ha hecho posible, en gran medida, la explosión en el crecimiento de Internet y la World Wide Web. Mientras mayor sea el "mercado" del conocimiento codificado, mayor (en principio en todos los casos) beneficio de la codificación. La consecuencia es que cualquier pieza del conocimiento codificado tiene muchos más potenciales usuarios. Y si el potencial se realiza, crecen los beneficios netos sociales que nacen de la codificación.

4.6 Los riesgos de la codificación.

4.6.1 Exceso de definición.

El principal riesgo objeto de numerosos debates es el riesgo de la sobre definición. Detallar los procesos y describir pasos y procedimientos tiene un umbral de eficiencia a partir del cual se cae en la sobre especificación y se anula la capacidad descriptiva de la documentación. En términos de la gestión del conocimiento, el proceso de documentación es un proceso de codificación del conocimiento y en esa actividad, Villavicencio y Salinas (2002) recalcan que los abusos de codificación pueden reducir los espacios de aprendizaje y provocar a la larga un estancamiento en la evolución de la organización.

Debido a su importancia en el sistema de gestión del conocimiento, fundamentalmente en empresas tecnológicas, estas cuestiones se desarrollarán en el capítulo 6.

4.6.2 Dependencia tecnológica

Con las tecnologías de la información no se están guardando documentos, sino una serie de instrucciones que es necesario interpretar y manejar por medio del equipo y los programas adecuados. Así que con cualquier distracción sobre los componentes complementarios de un sistema de conocimiento codificado (continuidad de lenguajes, guardar los programas que permiten el acceso a archivos anteriores) se corre el riesgo de alterar de manera irremediable la memoria global de la sociedad.

4.6.3 Almacenamiento de la codificación

Dado el crecimiento exponencial de todo tipo de documentos, ¿en verdad necesita guardarse todo? De no ser así, ¿Qué se guarda? ¿En qué medio, electrónico o impreso?

Los costos unitarios del almacenamiento y la recuperación de datos a corto plazo pueden haber disminuido, pero todavía hay problemas importantes respecto a memorizar, archivar y mantener el acceso para los documentos viejos. Los nuevos medios electrónicos de almacenamiento no son tan estables; de hecho, son inestables en comparación con el papel de celulosa bajo en sulfito en el que se han impreso los buenos libros por muchos años. Además, los lenguajes artificiales utilizados en la codificación de la información para su procesamiento en computadora también son menos estables en comparación, ya que están más propensos a volverse obsoletos de pronto, lo que requeriría que el corpus de la información almacenada “emigrara” periódicamente a un nuevo código que los nuevos programas sean capaces de leer.

En la era digital, el almacenamiento de la información se ha convertido en un asunto menos relacionado con archivar que con un proceso de renovación continua, una empresa cultural para la que las sociedades alfabetizadas resultan no estar bien preparadas.

4.6.4 Dificultades en la implantación.

Para evitar el riesgo de que no se implante con éxito un proceso de codificación del conocimiento Staples et al. (2001) advierte que es conveniente tener presente los problemas más comunes que se pueden presentar a la hora de abordar esta actividad:

- Por una parte, el conocimiento codificado no es necesariamente el más valioso para la empresa. Así, los esfuerzos para codificar conocimiento pueden centrarse en el conocimiento que es más fácil de capturar pero no necesariamente sobre el que tiene un mayor potencial a largo plazo para sostener el éxito de la empresa.
- La conversión de experiencias poco estructuradas en algo que puede ser almacenado y recuperado puede hacer que el conocimiento pierda su verdadero valor.
- La falta de preocupación para codificar el conocimiento tácito codificable, lo cual está relacionado con la falta de comprensión de la importancia de la codificación, la falta de rutinas y la poca claridad sobre cuál es el conocimiento tácito codificable.
- La dificultad para transformar las rutinas existentes y adoptar nuevas prácticas de compartir y codificar conocimiento. Además, tal y como hemos comentado en un apartado anterior, la inestabilidad de la estrategia tecnológica es un factor que afecta la generación de rutinas para codificar el conocimiento tácito codificable.

4.6.5 Costes

Otro riesgo importante a tener en cuenta es el coste que la codificación del conocimiento supondrá a la empresa, ya que este coste puede llegar a ser considerable y es vital hacer una buena estimación con antelación. Si la empresa se quedara sin recursos económicos, el riesgo de que la codificación no haya cumplido las expectativas es más que evidente.

A pesar de las ventajas del proceso de codificación, también hay que tener en cuenta que este proceso implica una serie de costes derivados tanto, de la posibilidad de facilitar la transferencia involuntaria de conocimiento a los competidores, como de la creación y mantenimiento de las infraestructuras y procesos derivados de la codificación.

Sin duda alguna las ventajas de la codificación no son gratuitas. La codificación del conocimiento supone costes específicos. Cohen y Bacdayan, (1994) apuntan que los costes directos incluyen el tiempo, los recursos y la atención directiva que se ha de invertir en el desarrollo y la actualización de los instrumentos específicos de trabajo, mientras los costes indirectos incluyen un posible aumento del coste de "errar el tiro" o la aplicación inadecuada de la rutina si la codificación se realiza incorrectamente. Otro

coste indirecto es consecuencia del aumento general de tareas de apoyo dedicadas a la formalización y estructuración de la ejecución de las tareas. La preocupación por estos costes es legítima e histórica. Un largo debate originado por el trabajo de Max Weber (1922), ha ocupado a teóricos organizativos sobre las ventajas y desventajas de la formalización, un fenómeno de naturaleza similar a la codificación de conocimiento.

4.7 Análisis Coste Beneficio de la codificación.

Evaluar los costes y beneficios de la codificación puede ser una tarea muy compleja. Y esto es debido efectivamente por la interacción de tres de sus aspectos. Cowan et al. (2000) indican que para entender el origen y la magnitud de los costes y beneficios, es necesario abordarlos desde el contexto del entorno del conocimiento.

Un primer y sencillo argumento es que los incentivos dependerán en gran medida de las posibilidades de los procedimientos de codificación sobre la base de lenguajes y modelos preexistentes. En esta clase de situación, los costes fijos generalmente han disminuido: lenguajes y modelos han sido desarrollados por trabajos pasados y son conocidos por codificadores y usuarios. El único coste es entonces el coste variable. Cowan et al. (2000) argumentan que, de esta manera, es útil diferenciar entre contextos de estabilidad y contextos de cambio. La definición de estabilidad será muy simple: estabilidad describe situaciones donde la codificación pasa por la base de lenguajes existentes y modelos existentes.

Por supuesto que la estabilidad no es permanente. El que los lenguajes existentes y modelos usados por los agentes sean suficientes para pasar al conocimiento codificado no implica que no haya cambios en el conocimiento o en el lenguaje. Hay muchos ejemplos de contextos de innovaciones estables, como los de Hummon (1984) y Steinmueller (1991), donde los cambios tecnológicos actúan como un procedimiento estándar, a lo largo de trayectorias bien definidas y predecibles.

4.7.1 Costes y beneficios en un contexto estable

Cowan et al. (2000) afirman que en un contexto estable (cuando hay un conjunto de personas que han hecho una inversión inicial para desarrollar un lenguaje y para mantener procedimientos eficientes para la adquisición del lenguaje a nuevos participantes) la transferencia de mensajes puede asimilarse a la transferencia del conocimiento, y el almacenamiento de mensajes significa la grabación del conocimiento.

En el lado de los beneficios, la eficiencia en la codificación será mayor en los grandes sistemas que tengan requerimientos específicos respecto a la coordinación entre los agentes. Cowan y Foray (1997) identifican cinco clases de situaciones:

- Sistemas que incluyen muchos agentes y muchas localizaciones. Es obvio que los beneficios estarán positivamente relacionados con la cantidad de usuarios potenciales. Mientras más agentes a los que difundir la información, mayor es el beneficio. De esta manera, la codificación proporcionará altos beneficios en

sistemas estables caracterizados por requerimientos específicos de transmisión del conocimiento y la comunicación. Esta clase de requerimientos resultan de las tendencias de deslocalización y externalización o del desarrollo de la investigación conjunta, implicando la distribución espacial de las actividades en muchos lugares. Este primer efecto puede ser apreciado sin ninguna ambigüedad, por ejemplo el que considera Mulligan (1994) en el ámbito de la ciencia: “un humilde hp 9000 altera radicalmente la manera en la que se realiza las investigaciones científicas por la gran cantidad de intercambio de la información: cada día, cerca de 20000 correos electrónicos difundidos a más de 60 países es la teoría de los nuevos papeles académicos, los lectores pueden desarrollar a través del beneficio del acceso a los artículos completos. Y cada día, cerca de 45000 físicos de todo el mundo acceden a los archivos electrónicos para encontrar o añadir nuevos asuntos”. Este efecto se produce con sólo un “clic” o una red – es una comunidad que comparte códigos comunes y conocimiento tácito para interpretarlos.

- Sistemas fuertemente basados en la recombinación y reutilización y que tienen ventaja en la no acumulación del conocimiento existente (más que en la innovación independiente). En sistemas (estables) de innovación donde los avances y novedades proceden, sobretodo, de la recombinación, la reutilización y la no-acumulación, los beneficios de la codificación son importantes. Gibbs (1994) afirma que los pequeños progresos en la productividad de la ingeniería del software son debidos a la excesiva dependencia de las técnicas que utilizan (comparadas por ejemplo con la ingeniería química). El esquema que Gibbs tiene en mente es que una vez que un algoritmo se escribe como parte de un código, puede, en principio al menos porque existe en una forma codificada, ser usado en muchas aplicaciones. La dificultad de hacerlo se plantea en parte por la falta de estandarización tanto en la forma en la que el código se escribe como en la forma en la que el algoritmo se emplea. Esta falta de racionalización tecnológica de la codificación impide la completa realización de las oportunidades que ofrece el modelo de la reutilización y la recombinación.
- Sistemas que necesitan memoria. En tercer lugar, los sistemas que necesitan memoria (p.ej. empresas con ciclos largos de desarrollo, tasas altas de rotación de empleados, con problemas de dispersión geográfica de la plantilla o instituciones enfrentadas a una gran bifurcación tecnológica) serán ampliamente beneficiados por la codificación. En esos sistemas, demasiadas codificaciones pequeñas aumenta el riesgo de una “no-inención accidental”. MacKenzie y Spinardi (1995) muestran, por ejemplo, que el específico, local y tácito conocimiento en el diseño del arma nuclear fue tan importante como lo fue el riesgo de perder conocimiento después de pasar una generación de científicos e ingenieros. El riesgo de una no-inención (o no-innovación) accidental en la que existe mucho conocimiento tácito se pierde. No hay razón para no extender este argumento a campos donde se desarrollan importantes invenciones para el bien social (por ejemplo la cura del cáncer); en los que el riesgo de una no-inención accidental debe ser eliminado gracias a la codificación.
- Sistemas que necesitan una particular descripción de qué (y cómo) hacen los agentes. Los sistemas que tienen la específica necesidad de describir que hacen

los agentes (ya sea encontrar restricciones en los estándares de calidad, o patentar innovaciones, o entablar cualquier clase de relaciones contractuales con un socio) tendrán grandes beneficios de la codificación. Aquí también podemos incluir los sistemas enfrentados a transacciones comerciales ineficaces; dónde Gunby (1996) afirma que los mecanismos tradicionales de garantías legales, seguros, reputación y pruebas no son eficaces para mitigar los efectos de la información asimétrica. Las prácticas de grabar la producción, que es una forma de codificación basada en las recientes innovaciones del lenguaje (creando estándares para archivar datos) tienen como objetivo reducir esas asimetrías.

- Sistemas caracterizados por un intenso uso de la tecnología de la información. Por último, pero no por ello menos importante, una serie de situaciones se ocupan de la falta de productividad que surge del uso de las TICs, debido a una incompleta codificación. Cowan (1995) indica que aprovechar por completo el potencial de la productividad que surge de las TICs requiere no solo la adopción de tecnología sino también cambios organizativos. Pero una empresa bajo cambios organizativos no quiere perder funcionalidad en el proceso. La empresa debe desarrollar conjuntamente las nuevas tecnologías y las estructuras organizativas para reproducir viejas funciones y crear otras nuevas. Es obvio que si muchas de las viejas funcionalidades residen en el conocimiento tácito, o dependen fuertemente en ellos, este aspecto será extremadamente difícil. Baumol et al. (1989) comentan que cuando la presencia de conocimiento tácito opera como un “cuello de botella”, impidiendo la completa realización del potencial de productividad, la empresa puede esperar grandes beneficios de la codificación.

En todos esos casos, donde se requieren operaciones importantes de transferencia, recombinación, memorización, descripción y adaptación (a las TICs) sobre el conocimiento existente, sería muy costoso e ineficiente mantener el conocimiento tácito. Describiremos estas situaciones como “exceso de conocimiento tácito”.

Una dimensión que puede ser conflictiva con estos beneficios tiene que ver con el problema de la proliferación de códigos locales en los niveles tanto inter- como intra-organizacional. Esta tendencia crea una alta incompatibilidad de costes e incrementa los costes de recodificación, que pueden ser considerados como una importante (y descuidada) dimensión de la paradoja de la productividad.

Otra dimensión conflictiva tiene que ver con la coexistencia de lenguajes, creados con el propósito de codificar el mismo tipo de conocimiento. La referencia aquí es de nuevo la incompatibilidad de costes. Podemos esperar un decremento sustancial de los beneficios si sobreviven varios dialectos. Este el caso, por supuesto de los lenguajes naturales (español, inglés, francés,...). Pero también es el caso de lenguajes especializados, tales como apunta Belofsky (1991) en el ejemplo de las dos “proyecciones multivistas ortográficas” (conocidas como “proyección del primer ángulo” y la “proyección del tercer ángulo”), que facilita la transferencia de conocimiento desde los ingenieros y arquitectos a los mecánicos y obreros.

Está claro que el clásico compromiso entre diversidad y estandarización que se desarrolla en la literatura de las opciones tecnológicas aplica a este caso. La existencia

de más de un lenguaje conlleva la posibilidad de la pérdida de acudir a redes externalizadas y un posible exceso irreversible de diversidad. David (1987) argumenta que se puede explicar mediante el uso de las mismas técnicas desarrolladas en las economías de los interfaces estándares para analizar las redes de información y las cuestiones de compatibilidad de las partes constituyentes de los sistemas de información.

En esta misma línea, cualquier estrategia usada en el caso de exceso de diversidad tecnológica puede extenderse a este problema: implementar anticipadamente lenguajes estándares, construir un meta-lenguaje y crear traductores, son valiosas estrategias para forzar la convergencia o al menos decrementar el coste asociado a la diversidad.

4.7.2 Costes y beneficios en el contexto de cambios.

Las situaciones de estabilidad en el entorno del conocimiento no son universales. Es más, es frecuente el caso en el que el entorno de conocimiento exhibe cambios constantes. Los modelos y los lenguajes son fluidos, y la comunidad de agentes que trabajan con los modelos y lenguajes están constantemente cambiando.

La fluidez de los lenguajes implica que existe una cierta cantidad de incertidumbre sobre lo que los mensajes actualmente significan porque hay incertidumbre, y quizá cambio, respecto al vocabulario en el que están escritos. Incluso cuando, por ejemplo, los artículos científicos expresan conocimiento en un lenguaje natural habrá cuestiones de jerga, definiciones de palabras que son específicas a la comunidad, y vocabulario que es específico al modelo usado. Estos sólo pueden ser aprendidos tal y como el modelo en sí llegó a ser entendido.

Cowan y Foray (1997) afirman que la codificación tiene lugar en este contexto. Los documentos científicos se escriben sobre nuevos campos. Pero los beneficios de la codificación llegan en gran medida del modelado y desarrollo de lenguajes. Puede haber una competencia entre los diferentes modelos básicos, y entre los principios básicos y vocabulario del lenguaje. Hasta que esta competición se resuelva, las cuestiones respecto a la difusión discutida en los párrafos previos será problemática. La comunidad de potenciales generadores y usuarios de conocimiento tendrán dificultades en la comunicación, y el valor de la codificación del conocimiento que surge de la diseminación será reducida. De esta manera, la codificación en este entorno tiene algún valor derivado del contenido del mensaje, pero tiene más valor como inversión (una contribución a la resolución de la competición entre lenguajes).

Es en este contexto de cambios donde podemos encontrar situaciones de exceso de codificación. Es decir, la acumulación de generaciones sucesivas de códigos puede prevenir el desarrollo de conocimiento radicalmente nuevo, simplemente porque explicarlo y entenderlo requeriría códigos enteramente nuevos.

Arrow (1974) argumenta que la codificación supone rigidez y uniformidad organizacional a medida que la comunicación y eficiencia de la transacción se incrementa. La necesidad de códigos mutuamente entendibles en una organización impone que se requiera una uniformidad en el comportamiento de los participantes.

Ellos se especializan en la información que puede ser transmitida mediante códigos, así que aprenden más en la dirección de su actividad y se vuelven menos eficientes en adquirir y transmitir información que no se ajusta fácilmente a los códigos.

El Segundo problema, entonces, tiene que ver con el exceso de inercia. Existen altos costes fijos que nacen en este proceso de codificación en un contexto de cambios. Conciernen en particular el desarrollo y aprendizaje del lenguaje compuesto por los nuevos códigos. Aproximadamente, los costes nacen durante el periodo en el que el entorno de conocimiento está fluyendo, mientras que los beneficios se reciben durante un periodo de estabilidad.

Durante un periodo de cambio, la infraestructura se desarrolla, los lenguajes y modelos se construyen, aprenden y estandarizan, y crece una comunidad de agentes que comparten conocimiento tácito. Todas estas inversiones contribuyen a la reducción en la fluidez del entorno, y conduce a la estabilidad. Cuando se alcanza la estabilidad, los beneficios de la codificación pueden ser significativos. Los lenguajes y modelos existen, lo que hace que se elimine uno de los costes. Y la red de posibles usuarios de conocimiento existe y es relativamente grande, lo que incrementa los potenciales beneficios.

Pero si el desarrollo de nuevos lenguajes y modelos destinan los costes fijos a una generación mientras muchas futuras generaciones se benefician de la nueva infraestructura de codificación del conocimiento, hay un problema intergeneracional que tiene que ver con la falta de incentivos privados para el desarrollo de lenguajes.

Konrad y Thum (1993) indican que las soluciones que pueden mitigar este tipo de problema de inconsistencia temporal podrían ser:

- El desarrollo de mercados relevantes (que pueden incrementar significativamente los beneficios incluso de la primera generación de desarrolladores)
- Con la creación de algún tipo de institución permanente para que no se produzca una discontinuidad abrupta en el futuro.
- El cultivo de preferencias altruistas para el bienestar de las generaciones venideras, a través de las cuales pueda ser legado un mayor stock de conocimiento útil.

4.8 La codificación lingüística del conocimiento

Camara (2004) expone que a través del arte y la ciencia se manifiesta y se representa el conocimiento:

- La ciencia utiliza, sobre todo, el lenguaje como mediador del conocimiento estructurándolo y representándolo, y la tecnología desarrolla artefactos que son productos derivados de conceptualizaciones cognitivo-lingüísticas. Los desarrollos tecnológicos son aparatos o máquinas resultantes de pensamientos

organizados con un objetivo concreto. En este sentido, se podría decir, que, en parte, cualquier máquina es una aplicación de conocimiento especializado y, por tanto, una representación aplicada del conocimiento.

- El arte es también otra de las manifestaciones para la representación de conocimiento. Sin embargo, la transferencia de sus contenidos es mucho más difícil de consensuar. La razón de esta dificultad estriba en el papel que desempeña la percepción, es decir, el conocimiento sensorial o emocional, con respecto al peso que adquiere el conocimiento cognitivo o intelectual. Esto supone que los mecanismos argumentativos para defender el significado que contiene la obra representada, de alguna manera, estén sesgados igualmente por la percepción del agente que interpreta la pieza representada. La materia prima utilizada para estos fines es muy diversa y no es patrimonio específico de un ámbito. Así, pueden encontrarse recursos lingüísticos, visuales y auditivos en cualquier manifestación de arte, ciencia y tecnología.

Dado el carácter científico de este trabajo, nos concentrarnos en la representación del conocimiento científico.

Dentro de los recursos que se pueden utilizar para la representación de conocimiento, destaca el papel del lenguaje por su relevancia particular. El lenguaje es básico en el desarrollo y transferencia de contenidos. Kandel et al (1995) han demostrado que las tareas vinculadas a la gestión del conocimiento están íntimamente relacionadas con los procesos cognitivo-lingüísticos que tienen lugar en el cerebro, y se materializan en forma de conocimiento representado a través del lenguaje.

El lenguaje es el reflejo del modelo de procesamiento del pensamiento humano. Y, visto desde esta perspectiva, el lenguaje empieza a entenderse y a explicarse más allá de las limitaciones impuestas por la lingüística tradicional, marcada por el entendimiento del lenguaje como fenómeno lineal.

El lenguaje es un fenómeno complejo y dinámico en el sentido de que está siempre en movimiento: por ser una herramienta universalmente utilizada que siempre está siendo actualizada por los agentes que intervienen e interactúan en el acto de la comunicación, y porque produce manifestaciones de formas espontáneas dependientes de las condiciones psicológicas, fisiológicas y sociales. Steels (2000) indica que esta diversidad compleja siempre creciente debe buscar una coherencia dentro de la misma estructura del sistema y dentro de un entramado aparentemente caótico.

Esta concepción del lenguaje responde a un paradigma complejo como un mapa abierto, conectable, descentralizado y con múltiples jerarquías variables, reflejo de las estructuras y la organización del conocimiento humano (bioconocimiento). Se trata de un paradigma de conocimiento complejo alternativo al paradigma lineal de los modelos inductivos y deductivos (causa-efecto).

4.8.1 Recursos lingüísticos

Por recursos lingüísticos se entiende el lenguaje natural y los términos cuya naturaleza de representación también es lingüística. Camara (2004) detecta cuatro formas en las que pueden aparecer los recursos lingüísticos representando extractos de conocimiento:

- La representación lingüística textual aglutina el lenguaje natural y los términos, que conjuntamente dan forma a cualquier tipo de documento presentado en forma de relato verbal, como el artículo, la guía, el informe, el manual o la tesis, entre otros. Este tipo de representación se considera de carácter analítico o explicativo, frente a las representaciones estructurales, que son la síntesis de un análisis.
- La representación lingüística sintética se abastece para la producción de sus contenidos, ante todo, de los términos estructurados para representar sistemas conceptuales (tesauros, ontologías, glosarios, etc). Así mismo, los términos pueden aparecer en diagramas y mapas, como único recurso lingüístico o en combinación con otros recursos de tipo visual o auditivo.
- Los documentos estructurados en forma de relato verbal pueden abastecerse de lenguaje natural y términos para la representación de sus contenidos y pueden utilizar otros recursos, como representaciones estructurales, abastecidas de imágenes, símbolos o iconos.
- Las representaciones multimedia son una combinación de materia prima y tipo de representación; en cuanto a la materia prima utilizada, pueden hacer uso de lenguaje natural, términos, imagen o sonido, y pueden aparecer en extractos de conocimiento de tipo analítico o de tipo sintético.

Galinsky (2001) afirma que lo que llamamos conocimiento son conceptualizaciones representadas que explican la naturaleza de nuestro mundo real y abstracto, pero el conocimiento en sí, no existe en la naturaleza. Se representa el contenido de los razonamientos, a través de símbolos de tipo lingüístico, de tipo visual o de tipo auditivo. Los símbolos son transportadores de conocimiento que pueden apelar a cualquiera de los sentidos para hacer llegar el contenido semántico y el sentido perceptivo que encierran.

Codina, (2001) argumenta que el lenguaje natural, la palabra y el texto, son la forma más potente de representación del conocimiento y, por tanto, de gestionar la información. Para Junyent (2003), la generación, representación y transferencia del conocimiento a través del lenguaje es uno de los factores determinantes para la supervivencia de nuestra especie sobre otros homínidos.

Los términos son unidades de naturaleza lingüística que representan conocimiento de un determinado dominio y son consensualizaciones para denominar conceptos a fin de poderlos transmitir. Cámara (2004) razona que, al ser el lenguaje un fenómeno determinante para el desarrollo de conocimiento, y unido a la funcionalidad de los términos, se puede afirmar que no existe conocimiento sin terminología.

Los términos son parte fundamental de la estructura textual que representa el conocimiento y se presentan tanto en la microestructura como en la macroestructura de cualquier representación lingüística de contenidos.

Cámara (2004) define microestructura como el conjunto de unidades que:

- componen la matriz cognitiva (sistema conceptual) sobre la que se construye un texto (documento);
- sirven de elementos identificadores, o descriptores para identificar el contenido de un documento;
- materializan las presentaciones estructurales (gráficos, esquemas, mapas mentales, tesauros, ontologías...) de conocimiento;
- integran los nodos, vértices de los diferentes esquemas de representación del conocimiento (mapas conceptuales, marcos, redes semánticas...) utilizados en el ámbito de la inteligencia artificial;

A su vez, Cámara (2004) define macroestructura como:

- los productos textuales finales: el texto adaptado a diferentes tipos de documentos: material didáctico, tesis, informes de proyecto, informes anuales, guía, manual, etc.;
- los sistemas de representación estructural para gestionar los productos textuales finales y para gestionar el conocimiento de una disciplina.

4.8.2 Disciplinas vinculadas a la codificación lingüística.

Los términos desempeñan un papel preferente en la representación del conocimiento. Si se considera cuál es su naturaleza (lingüística), lo que representan (conocimiento) y su función (comunicar y transferir conocimiento), no es difícil deducir qué disciplinas troncales intervienen en su estudio.

En la figura 4, Cámara (2004) representa de qué disciplinas se obtienen los fundamentos en los que se apoyan toda la investigación y el método relacionados con la representación lingüística del conocimiento (lingüística, ciencias del conocimiento, y ciencias de la información y la comunicación). Así mismo, en el esquema, se apuntan qué materias interdisciplinarias se han derivado de las fundamentales y dan respuestas a las incógnitas relacionadas con la representación de contenidos, en cuanto unidades que representan tanto la microestructura como la macroestructura que subyacen a cualquier representación lingüística de conocimiento.

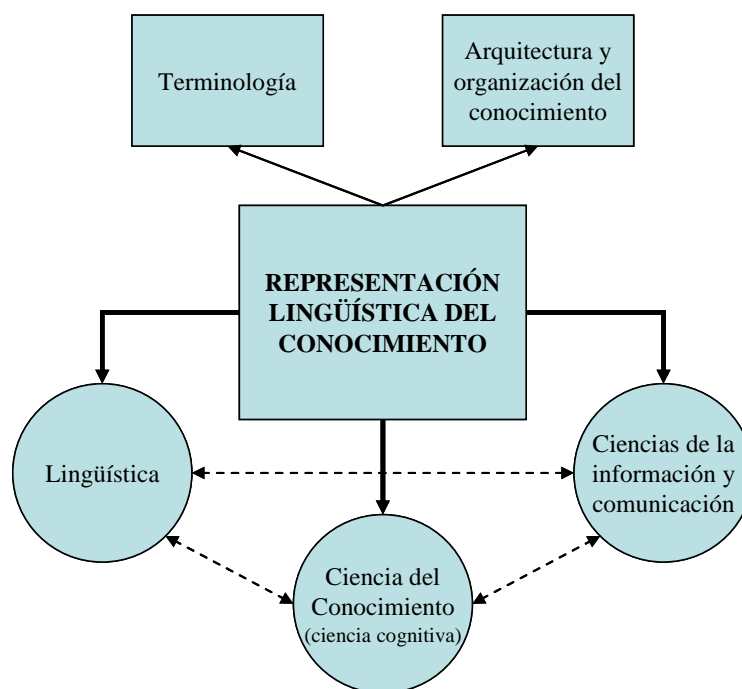


Figura 4: disciplinas vinculadas a la lingüística

La **terminología** es una ciencia interdisciplinar que se nutre de un conjunto específico de conocimientos conceptualizado en otras disciplinas (lingüística, ciencia del conocimiento y ciencias de la información y la comunicación). Así mismo, Cabré (1998) asevera que es una ciencia transdisciplinar porque los productos terminológicos son las piezas de representación lingüística en las que debe apoyarse cualquier campo de conocimiento científico para poder adquirir, generar y transferir el conocimiento específico de cualquier dominio. Eso quiere decir que en el campo del derecho, de la medicina, o de la física, por ejemplo, la disciplina de la terminología desempeña un papel fundamental como proveedora de términos transportadores de conocimiento que median en la comunicación, como identificadora de reglas subyacentes en la generación y relación de los términos, y como método y habilidad de trabajo.

Cabré (1993) recuerda que la práctica terminológica y su aplicación para conceptuar los avances científicos y técnicos data de hace tres siglos. Sin embargo, no será hasta 1931 cuando se publique el primer trabajo de sistematización metodológica para el desarrollo de técnicas terminológicas y terminográficas. Wüster (1979/1998) desarrolla planteamientos teóricos (Teoría General de la Terminología) que han servido de base según Cabré (1999) para fundamentar la terminología moderna, de la que ya se han derivado nuevas concepciones.

La **organización del conocimiento** es una disciplina de formación reciente que estudia las leyes, los principios y los procedimientos para estructurar el conocimiento especializado e identificar y establecer la arquitectura en la que se sustenta el conocimiento de cualquier dominio.

Mucho antes de que la organización del conocimiento alcance el estado de disciplina, es tratada por Aristóteles en su empeño por fragmentar el mundo para entenderlo y, así,

poder clasificarlo. Desde entonces, esta actividad ha sufrido cambios espectaculares. De concepciones puramente metafísicas y especulativas, desarrolladas en el terreno de la filosofía, la epistemología y la ontología, se ha pasado a concepciones físicas e informático-pragmáticas, desarrolladas desde diversas ramas (biblioteconomía, documentación, epistemología, informática, inteligencia artificial, semiótica, terminología, etc.), desplegadas a partir de las ciencias fundamentales: lingüística, ciencias del conocimiento y ciencias de la información y la comunicación.

Los inicios de esta disciplina como tal coinciden con los avances de la documentación y la biblioteconomía. Paul Otlet (1947) da pautas para generar instrumentos fundamentales para la gestión de conocimiento científico, basando toda su estrategia en la organización (clasificación) de los contenedores (documentos) donde se guarda el conocimiento representado, es decir, la información. Los trabajos del alemán Wilhelm Ostwald, premio Nobel de química en 1906, son también transcendentales en el desarrollo de la disciplina de la organización del conocimiento. Ostwald (1912) plantea por primera vez soluciones para la organización del trabajo intelectual (el trabajo intelectual es conceptualmente igual a lo que denominados conocimiento científico). Con su propuesta de que el conocimiento debe organizarse de forma fragmentada, estandarizada, normalizada e internacionalizada, Ostwald se convierte en el precursor de la web semántica.

Este campo de estudio es de naturaleza interdisciplinar y se nutre de los aportes recibidos de la lingüística, las ciencias del conocimiento y las ciencias de la información y la comunicación. El objeto de esta disciplina es, por una parte, la organización del conocimiento explícito (conocimiento representado), es decir, socializado o registrado a fin de optimizar la circulación del conocimiento en la sociedad y, por otra, el desarrollo de metodologías que sustenten paradigmas de conocimiento válidos, escalables y ampliables. García Marco (1998) considera que también es una disciplina científica, esto es, sujeta al método científico, de carácter aplicado.

En la siguiente figura, Camara (2004) representa el despliegue de disciplinas que, desde las ciencias fundamentales, pueden vincularse al ámbito de la representación del conocimiento, desde la perspectiva de su microestructura (terminología) y de su macroestructura (representación del conocimiento).

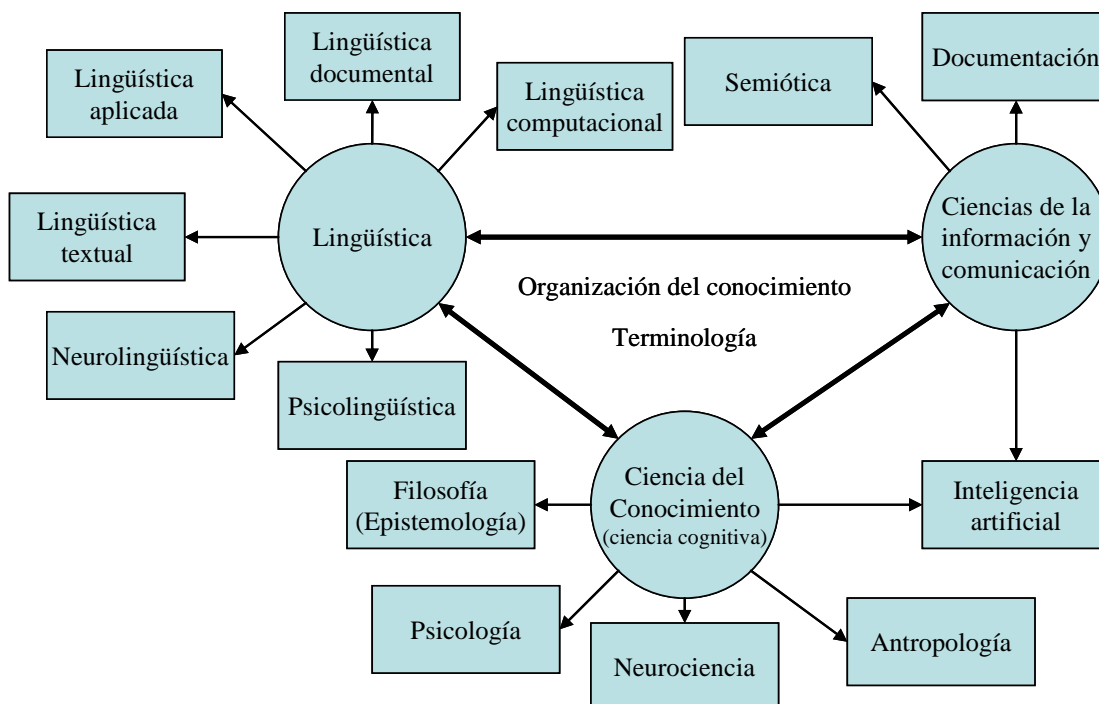


Figura 5: disciplinas vinculadas a la codificación del conocimiento

La esfera que representa las disciplinas de la terminología y la organización del conocimiento tiene flechas en su interior en varias direcciones. Con ellas, se representa que estas materias están vinculadas a todas las disciplinas derivadas de las ciencias troncales resaltadas en círculos mayores. El mapa da cuenta de la interdisciplinariedad y complejidad del ámbito en el que se desarrolla la ingeniería lingüística.

Si partimos de la propuesta de Cámara (2000), derivada de una propuesta clasificatoria de García Gutiérrez (1984 y 2002), el mapa muestra las disciplinas de las que se parte para el estudio de la representación, organización y aplicación del conocimiento (lingüística documental y ciencias del conocimiento). Esta clasificación es una prueba más de las diferentes formas de ordenar el saber, y de disponer de etiquetas que representen nuestras categorías con fines específicos y especializados.

4.8.3 Relevancia de la codificación en la ingeniería lingüística

Desde que la sociedad de la información global y multilingüe (caracterizada por la omnipotente influencia de las tecnologías de la información y de la comunicación) ha expandido su influencia vertiginosamente, la gestión de los recursos lingüísticos se ha convertido en una necesidad ineludible para la mayoría de las organizaciones públicas y privadas.

La industria del conocimiento irrumpe con fuerza ofreciendo respuestas que acaben con la “infotoxicación” y que contribuyan a la creación de más conocimiento a partir de información codificada como simples datos. Esta industria, por su estrecha relación con la codificación del conocimiento como información textual, depende de métodos,

herramientas y recursos lingüísticos que han venido desarrollando las industrias de la lengua.

Camara (2004) dice que cada año se crean nuevas asociaciones, comunidades y grupos de interés que se involucran como usuarios o como desarrolladores en actividades relacionadas con la gestión de recursos lingüísticos. Para abastecer de materia prima al mercado son necesarios nuevos lenguajes relacionados con dominios de especial interés, relacionados también con los diferentes registros de la lengua, con una dimensión diferente entre el lenguaje escrito y el oral (chat), y por último, relacionados con la tipología y estructuración textual. De esta materia prima lingüística se abastecen ámbitos tan diversos como el de la traducción e interpretación, el reconocimiento de voz, la recuperación de información en entornos abiertos (Internet) y cerrados (Intranet), y la gestión del conocimiento. Todas estas aplicaciones a las que se dedica la ingeniería lingüística están relacionadas con la gestión de la información y del conocimiento. Son aplicaciones que precisan trabajar con el conocimiento representado en forma de productos terminográficos, sistemas conceptuales y otros recursos lingüísticos, así como métodos, herramientas y experiencia acumulada en este sector de la industria de la lengua.

Desde el punto de vista cognitivo, los sistemas desarrollados en el ámbito de la ingeniería lingüística están estrechamente vinculados a la ingeniería del conocimiento, porque sus planteamientos se basan en estructuras lingüísticas que conforman sistemas conceptuales variables que ofrecen respuestas probables. Esta visión cognitiva se está aplicando en muchos de los retos asociados al procesamiento del lenguaje natural. En este sentido, Hutchins y Somers, (1995) comentan que los nuevos sistemas de traducción automatizada, por ejemplo, se basan en corpus textuales paralelos (memorias de traducción) que se indexan utilizando bases de conocimiento que ayudan a estructurar y a eliminar la ambigüedad de los posibles resultados que ofrece la máquina.

Es un planteamiento diferente con respecto a las primeras épocas de la ingeniería lingüística. La primera época parecía ser meramente estadística hasta que se introdujeron estrategias basadas en conocimiento lingüístico de tipo morfológico y después sintáctico. El renacimiento de la relevancia de la semántica en los desarrollos de la ingeniería lingüística es el resultado de los avances aportados por la ciencia del conocimiento. Esta evolución de planteamientos corresponde a un cambio de paradigma del conocimiento, de uno lineal a otro complejo.

Los sistemas de recuperación de información también han sufrido una transformación cuya evolución también ha repercutido en su denominación: sistemas de recuperación de conocimiento, sistemas de gestión de conocimiento. De Mey (1980) ha tipificado cuatro etapas en el desarrollo del tratamiento de la información que coinciden las que determinan la evolución de la ingeniería lingüística.

La ingeniería lingüística es una actividad eminentemente aplicada. Es el resultado de representar en forma de artefactos y lenguajes artificiales las deducciones que se obtienen del área de trabajo de la lingüística textual, de la lingüística computacional, de la informática, de la terminología y de la organización del conocimiento.

Cámara (2004) indica que las áreas de la ingeniería lingüística vinculadas al diseño y desarrollo de sistemas basados en conocimiento son:

- Adquisición de conocimiento (“knowledge acquisition”)
 - Identificación de unidades o grupos de conocimiento
 - Extracción de datos que representan unidades o grupos de conocimiento
- Modelado de conocimiento (“knowledge modelling”)
 - Desarrollo de metadatos interpretables e intercambiables de forma digital
 - Arquitectura del conocimiento (metodología para la macroestructura)
 - Macroestructura: Estructura para la clasificación, indización y recuperación del conocimiento (sistemas conceptuales)
- Representación de conocimiento (“knowledge representation”)
 - Creación de términos
 - Recopilación de unidades o grupos de conocimiento
 - Identificación de las relaciones semánticas entre unidades o grupos de conocimiento
 - Microestructura: estructuración de unidades o grupos de conocimiento
 - Macroestructura: estructura para la clasificación, indización y recuperación del conocimiento (sistemas conceptuales)
- Infraestructura para el desarrollo de ingeniería de conocimiento (“knowledge engineering development infrastructure”)
 - Bases de conocimiento (“knowledge base”)
 - Reglas de inferencia (“inference rules”)
 - Recuperación de conocimiento (“knowledge retrieval”)
 - Interacción y diseño de interfaces humano-máquina (“human-computer interaction /human-computer interface design”)

Todas estas áreas enumeradas de forma secuencial están relacionadas entre sí, como se intenta representar en la siguiente figura de Cámara (2004):

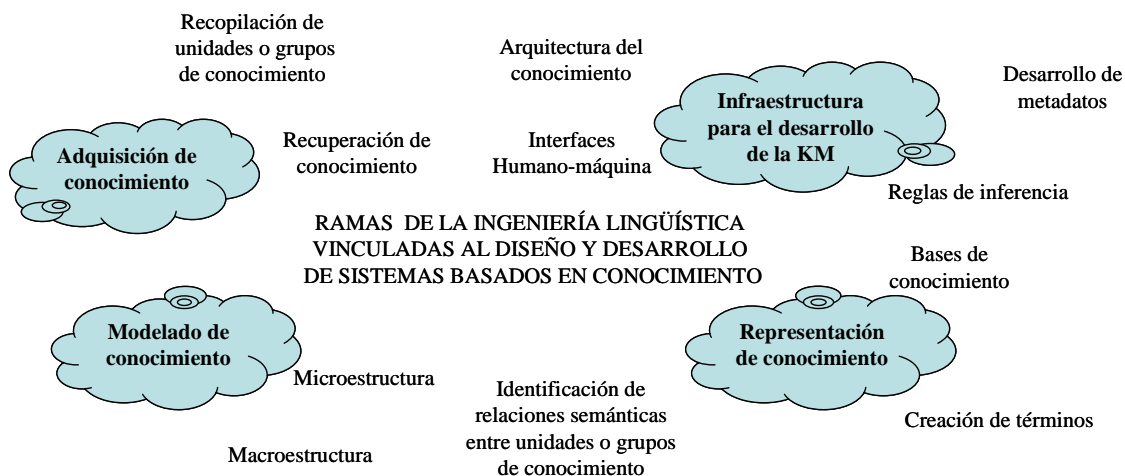


Figura 6: ramas de la ingeniería lingüística vinculadas a la gestión del conocimiento.

La adquisición de conocimiento está relacionada con una actividad humana que corresponde a la prospección de información (conocimiento) que se obtiene de un experto. En el ámbito de la ingeniería lingüística, y más específicamente de la ingeniería de conocimiento, también se desarrollan herramientas para la extracción automatizada de datos que representan conocimiento, por ejemplo, de tipo léxico o terminológico. Moreno (2000) indica que esta extracción se puede realizar de forma manual, a partir de diccionarios electrónicos y de corpórea textuales digitalizados.

La construcción y el mantenimiento de conocimiento sobre un dominio mediante diccionarios (enciclopedias), ontologías, clasificaciones (epistemografías), están vinculados a la representación del conocimiento. Ésta puede beneficiarse de la adquisición del conocimiento, influye en el modelado del conocimiento y, por último, es una pieza insustituible para definir infraestructuras de desarrollo para la ingeniería de conocimiento.

De las tareas correspondientes a la representación del conocimiento, Cámara (2000) expone que se deben desprender los parámetros vinculados a las propiedades:

- que ofrecen las estructuras semánticas descritas en el sistema conceptual decidido en el modelado del conocimiento, y
- pragmáticas generadas a partir del input de los usuarios, ya sea de forma directa, o bien mediante una interfaz de consulta y alimentación.

De estas propiedades se pueden derivar reglas para el desarrollo del motor de inferencia y de control, dos de los módulos que integran un sistema de gestión basado en conocimiento, como veremos más adelante. La forma de presentar el conocimiento, es decir, de publicarlo, dependerá de los criterios adoptados para la representación de la microestructura y de la macroestructura. Por tanto, la publicación del conocimiento está vinculado a su representación y modelado.

La importancia de la representación del conocimiento en redes multiasociativas basadas en probabilidades marca una etapa cognitiva que también repercute en la ingeniería

lingüística. Se trata de que los sistemas simulen la forma de procesar la información en el cerebro, lo que supone representar previamente esa forma de clasificar y estructurar el cerebro para alimentar el sistema con información que pueda generar instrumentos de gestión de conocimiento.

Para ello, la estructuración del conocimiento representado que conforma la llamada base de conocimiento de cualquier sistema experto es esencial. Eso quiere decir que cualquier proyecto de inteligencia artificial que utilice el conocimiento como clave de la resolución automatizada de problemas cognitivos deberá confeccionar en primera instancia un proyecto de representación de conocimiento de un ámbito concreto de especialidad para poder generar una buena base de conocimiento de donde se pueda inferir y transferir nuevo conocimiento.

4.9 Gestión documental

Green (2003) parte de las siguientes premisas:

- Toda organización de documentos que se base en las características de los documentos mismos constituye una codificación de conocimientos acerca de ellos.
- Las metodologías de organización documental son también metodologías para codificar el conocimiento acerca de los documentos (esta segunda premisa se deriva de la anterior).

Se habla de “la organización documental” en un sentido muy amplio; el argumento que se desarrolla engloba sistemas muy diversos que se utilizan para organizar y acceder a colecciones de documentos de todo tipo, desde los libros hasta los materiales gráficos y audiovisuales. Por “metodología” se entiende un conjunto de “métodos, principios y reglas” que se emplean para un propósito determinado; se supondrá que todo sistema concreto que se utilice para la organización documental (como el Dewey Decimal o las Reglas de Catalogación Anglo-Americanas) se encuentra sustentado por una metodología determinada. Este supuesto constituirá el punto de partida para cuestionar los principios en base a los cuales dichos sistemas funcionan.

4.9.1 Codificación según ubicación física

En este apartado se tratan el grupo metodologías que establecen una equivalencia entre la clasificación del documento y su ubicación física. La creación de esta equivalencia es un principio operante sobre todo en la organización de bibliotecas, que se apoyan en los sistemas de clasificación como el Library of Congress y el Dewey Decimal. Estos sistemas asignan a cada objeto una clave que sirve simultáneamente para describirlo y para señalar el lugar donde se almacena. Por ejemplo, en el sistema del Library of Congress, se asigna la clave “QE534.2.B64” al libro *Introducción a terremotos*, de Bruce Bolt. Esta clave contiene datos acerca del tema del libro, que en este caso se clasifica bajo las categorías “Ciencia”, “Geología” y “Geología” dinámica y estructural”; esta clasificación se registra con la primera parte de la clave, “QE534.2”.

También contiene información acerca del autor del libro, la cual se codifica como “B64” en la parte final de la misma.

Rosenberg (2003) expone que a veces a todo esto se anexan datos adicionales, como el año de publicación o un número de ejemplar para distinguir entre múltiples copias de un documento. El sistema de claves fue diseñado de manera que siempre habrá una sola manera de ordenarlas secuencialmente; y los objetos clasificados se ubican físicamente en el repositorio con base en esta secuencia. La misma clave simultáneamente describe el documento y determina dónde se coloca.

Los sistemas Library of Congress y el Dewey Decimal fueron creados en el siglo XIX pero han sido actualizados continuamente, sobre todo para adecuarlos a las necesidades cambiantes de descripción de obras. Se emplean en miles de bibliotecas en todo el mundo.

También se utilizan sistemas similares en archivos de materiales muy diversos. Por ejemplo, Magaña (2003) nos enseña cómo en la fototeca Pedro Guerra de la Facultad de Ciencias Antropológicas de la Universidad Autónoma de Yucatán emplea una “clave topográfica” que también funciona para describir las piezas y ordenarlas físicamente. El método que se usa para construir dicha clave se explica sencillamente a través de un ejemplo:

2A05_021, en donde la primera variable se encuentra representado por el número dos indicando el tipo de soporte de la pieza que en este caso corresponde a una placa seca o negativo cuya imagen se encuentra plasmada en cristal; la segunda variable cuyo lugar ocupa la letra A permite identificar el autor de la misma, siendo para este caso el señor Pedro Guerra Jordán; el lugar correspondiente a la tercera variable lo ocupa el numeral 05, mediante el cual proporciona datos acerca del tamaño del negativo, por tanto en el supuesto que nos ocupa indica el tamaño de una placa de 5 x 7 pulgadas; mientras que los últimos tres dígitos representados por la terminación 021 permite apreciar el seguimiento de las piezas por medio de un número progresivo, de acuerdo al formato o tamaño de las placas en cuestión.

4.9.2 Soportes de registros

Existe una relación muy estrecha entre la codificación de la información documental y los soportes o medios que la registran, que son:

- las etiquetas que se colocan en las espaldas de los libros y otros documentos.
- el orden físico de los documentos.

Las posibilidades y limitaciones de los soportes han influenciado sustancialmente la manera en que estos sistemas representan la información. Green (2003) considera que las limitaciones más importantes han sido

- el tamaño reducido de las etiquetas, el cual obliga a condensar todos los datos en una clave muy breve.

- la necesidad de ordenar todos los objetos en una sola secuencia.

Otro tipo de organización documental que han empleado tanto los archivos como las bibliotecas son los catálogos en papel y cartón, con fichas acerca de cada documento, las cuales casi siempre registran una mayor cantidad de información que las claves que se han descrito anteriormente. Los sistemas de tarjetas que se usaban (y en muchos casos, que todavía se usan) en las bibliotecas son quizá el mecanismo más conocido de este tipo; en esas instituciones, los tarjeteros permitían ubicar un libro a partir de su título, autor o tema, y proporcionaban descripciones de los materiales elaboradas con base en normas establecidas.

La situación en los archivos visuales, audiovisuales y sonoros ha ido mucho más variada. A menudo se siguen usando catálogos en tarjetas o en cuadernos, y a veces ni eso hay; y es menor el grado de normalización en el formato de las fichas.

Un ejemplo de una metodología de este tipo es la que propone la norma de catalogación de materiales sonoros editada por la IASA (Asociación Internacional de Archivos Sonoros) en 1998. Las reglas de catalogación de la IASA especifican los requisitos para la descripción e identificación de grabaciones sonoras y materiales audiovisuales relacionados, fijan un orden para los elementos de la descripción y especifican un sistema de puntuación para esa descripción. Fueron diseñadas para ser empleadas en archivos sonoros y audiovisuales como una guía para elaborar registros catalográficos y como norma para intercambiar información bibliográfica acerca de grabaciones sonoras y materiales audiovisuales relacionados.

Como se dice el texto citado, esa norma incluye reglas para el uso de signos de puntuación en las fichas. Éstos sirven para indicar el significado de los distintos componentes de la ficha. Por ejemplo, en la primera sección de ésta, se utiliza un diagonal ("/") para señalar la primera persona u organización responsable por la creación de una obra; para los demás autores o responsables se utiliza el punto y coma (";"). Para indicar una traducción del título principal, se emplea un signo de igual ("="). Y así sucesivamente; el sistema de puntuación constituye un eje fundamental para interpretar los elementos de un registro elaborado con base en esta norma.

¿Por qué se utiliza la puntuación así? ¿Qué ventaja tiene? ¿Por qué no escribir el nombre completo de cada elemento en la ficha? Por ejemplo, según esa norma una ficha podría empezar así:

Moonlight Shadow [grabación sonora] / Mike Oldfield.

Pero ¿no sería más claro así?

Título: Moonlight Shadow
Descripción del material: grabación sonora
Autor: Mike Oldfield.

Green (2003) comenta que uno de los motivos por los cuales ha resultado útil emplear la puntuación de esa manera (tanto en la norma de la IASA como en otras similares) es simplemente que así se ahorra espacio y por lo tanto se logra registrar una mayor cantidad de información en la superficie limitada de una tarjeta.

Otra huella de las limitaciones del soporte es el uso de abreviaciones, como "ed." para "edición" o "s.l." para "sin localizar", las cuales permiten condensar aún más los datos de la ficha. Y además de los problemas del espacio, hay otras limitaciones de este soporte que han repercutido en las maneras de codificar la información (por ejemplo la necesidad de generar un índice por cada tipo de punto de acceso, o de reunir en una misma tarjeta datos dispersos, para mayor comodidad del usuario).

Es necesario enfatizar acerca de las múltiples metodologías disponibles para la elaboración de catálogos en cartón o papel. La manera en que éstas representan el conocimiento documental también ha sido fuertemente influenciada por las características de los soportes físicos que emplean. Otra vez, el limitante fundamental ha sido la necesidad de inscribir en un soporte pequeño mucha información variada.

4.9.3 El uso de la tecnología

Se debe considerar también las metodologías que se usan para organizar un repositorio y codificar la información documental por medio de las computadoras. Destaca que, si bien este medio es revolucionario en muchos sentidos, la mayoría de los sistemas que actualmente lo emplean son una traducción casi directa de las formas de documentación que utilizan soportes de papel y cartón. De hecho, éste es el propósito explícito del formato MARC, como su nombre indica: "Registro catalográfico legible por máquina" ("Machine Readable Cataloguing record"). El desarrollo del formato empezó en los años 60 cuando apenas se creaban los primeros catálogos automatizados. El MARC propone representar en el medio digital la misma información que se inscribe en las tarjetas de cartón.

Si bien es una norma muy completa, el MARC sigue reflejando esta herencia. En particular, agrupa y ordena los datos de una manera que sigue los patrones establecidos con los catálogos en cartón o papel.

Furie (2001) expone que el MARC tiene algunas limitaciones muy específicas que se derivan del formato de archivo digital que emplea, que es el soporte propiamente dicho. (Un ejemplo es el campo 130, que incluye un mecanismo para indicar qué elemento del título se utiliza para ubicar la ficha en una secuencia alfabética. Este mecanismo es útil para títulos como "La Jornada" o "El viaje al centro de la tierra", que se colocarían bajo "J" y "V" respectivamente, y no "L" o "E". Sin embargo, por las limitaciones del formato digital empleado, es imposible señalar para este fin una palabra que inicie después del noveno carácter del título.)

4.9.4 Necesidad de nuevos métodos

Todos los sistemas mencionados en apartados anteriores integran en un solo movimiento intelectual la búsqueda de soluciones a los problemas de:

- Codificación y estructuración del conocimiento acerca de los documentos.
- Comunicación de este conocimiento.
- Registro físico.

Sin embargo, para abordar estos problemas se requieren herramientas bastante diferentes; y si bien debemos combinar soluciones de estas tres áreas para un fin concreto (que es organizar y acceder a una colección de documentos), es deseable que las limitaciones que encontremos en cada una de ellas afecten lo menos posible al desarrollo de las demás.

Por lo tanto, para construir prácticas de codificación más eficaces en primer lugar es necesario separar estas áreas y no confundirlas, como hacen los sistemas que se han descrito.

Hoy existen varias propuestas de sistemas y metodologías de codificación que ponen en práctica esta separación, en mayor o menor grado. En vistas del potencial que dichas propuestas encierran para resolver problemas que antes parecían irresolubles, así como la manera en que los supuestos básicos que las sustentan se apartan de modos de pensamiento anteriores, Green (2003) sugiere que estos nuevos sistemas representan el inicio de un cambio de paradigma en la catalogación y enumera los elementos centrales de este cambio:

- Se hace una separación analítica entre los problemas de codificación y estructuración del conocimiento, comunicación de éste y registro físico.
- El primer paso en la generación de un catálogo es el análisis del conocimiento que se tiene acerca de los documentos y la elaboración de una estructura abstracta que permitirá representar este conocimiento. Esta estructura refleja los puntos de vista de sus autores acerca de las cualidades esenciales de las cosas que se describen y las relaciones que existen entre éstas. En varios ámbitos, las estructuras de este tipo se denominan "ontologías" (en informática, el significado de ontología según el The Free On-Line Dictionary of Computing es "una especificación explícita y formal acerca de cómo representar tanto los objetos, conceptos y demás entidades que se supone existen en un determinado campo de interés, como las relaciones que se presentan entre ellos")
- Esta estructura abstracta será el componente central del catálogo. Se podrán elaborar múltiples expresiones del conocimiento que la estructura contendrá, en diferentes lenguajes y soportes físicos, para fines de comunicación entre personas o máquinas; pero el formato abstracto siempre será considerado el "original" o "molde" del cual las demás expresiones se derivarán.

Se ha elaborado esta idea del paradigma emergente no precisamente a partir del contenido de las nuevas propuestas de sistemas de gestión documental, sino sobre todo con base en una aplicación rigurosa de los principios que están operando detrás de ellas.

Algunos ejemplos de dichas propuestas son: FRBR, Topics Maps, Tesauros y web semántica.

4.9.5 FRBR (Functional Requirements for Bibliographic Records)

El informe "Requisitos Funcionales para Registros Bibliográficos" (FRBR en inglés), de la Federación Internacional de Asociaciones de Bibliotecarios y Bibliotecas (IFLA en inglés). Publicado en 1998, este informe propone una estructura informacional abstracta conformada por "entidades" (como "obra", "expresión" y persona), relaciones entre éstas, y "atributos" que las describen (como "título", "fecha" y "tema"). Green (2003) considera que esta iniciativa está menos desarrollada que otras, sin embargo contiene elementos que las demás no contemplan. Hickey (2002) cree que el FRBR está teniendo un impacto profundo en cómo se perciben los datos bibliográficos y Weston (2001) lo relaciona con cambios en los fundamentos de la catalogación, al afirmar que la necesidad de reevaluar los fundamentos de la catalogación se hizo muy evidente a raíz de los trabajos teóricos que llevaron a la publicación del FRBR en 1998. Por un lado se pensó que había inconsistencias por el uso intenso del ISBD, norma basada en principios que poca relevancia tienen en el ambiente electrónico, ya que fueron definidos a partir de preocupaciones que se relacionan con las versiones impresas del catálogo nacional.

Una de las estructuras más interesantes que propone el FRBR es la que establece la relación entre una *obra* (que se define como una creación intelectual o artística distinta), una *expresión* (que es la realización de una *obra*), una *manifestación* (la concreción de una *expresión* u *obra* en un tipo determinado de soporte) y un *objeto* (un ejemplar de una manifestación). Por ejemplo: un guión de Shakespeare sería una *obra*; la versión de esa *obra* editada por Kenneth Muir, quien cotejó el texto con varias fuentes, sería una *expresión*; el trabajo de diseño gráfico que permite apreciar esa *expresión* en un libro sería una *manifestación*; y un ejemplar del libro que contiene esa manifestación sería un *objeto*.

4.9.6 Topics Maps

Topic Maps tiene su origen en el grupo de Davenport, un foro destinado a productores de libros electrónicos que surgió a principios de la década de los 90. En 1993 se propuso la creación de una norma cuyo principal objetivo fuera posibilitar la fusión de índices impresos. Posteriormente evolucionó hacia otras estructuras (como tesauros), hasta llegar a ser una herramienta considerada en la web para la organización, representación y gestión del conocimiento. La primera versión oficial del estándar ISO/IEC data del año 2000. Los Topic Maps, se desarrollaron durante la década de los '90 para describir estructuras de conocimiento y vincularlas con recursos de información. Como tales, Peppers (2000) argumenta que constituyen una tecnología muy capacitada para la

gestión del conocimiento. Conocida como “el GPS del universo de la información”, los topic maps también se ocupan de proporcionar nuevas y poderosas formas de navegar en grandes corporaciones interconectadas.

A pesar de que es posible representar estructuras muy complejas usando topic maps, los conceptos básicos del modelo, Temas, Asociaciones y Ocurrencias (TAO) se comprenden muy fácilmente.

Las ventajas que tienen los topic maps para representar la web semántica se pueden resumir en:

- Los topic maps, pueden dar semántica a elementos que están en la web al organizarlos y describirlos, pero sin modificarlos.
- Perfiles de usuarios: los topic maps permiten adaptarse a distintas comunidades compartiendo recursos informativos.
- Navegabilidad e inferencia mediante estructuras semánticas. Lo cual mejora no solo la recuperación de información, sino también la gestión del conocimiento y el mantenimiento de los topic maps. En este punto es también interesante recordar la independencia de los recursos informativos que tiene topic maps, pudiendo navegar por nodos que no tengan ningún recurso asociado.
- Fusión con otras estructuras de conocimiento, permitiendo una gestión descentralizada.
- Buena escalabilidad y compatibilidad para adaptarse al creciente número de recursos de información.

Las posibles desventajas residen más en la poca madurez del estándar que en otros aspectos. Así se han revisado distintos aspectos, como:

- La necesidad de mejorar la capacidad de integrar las propiedades de los diferentes tipos de asociaciones dentro del esquema de definición de topic maps.
- La línea de investigación emergente sobre la forma de asegurar la consistencia de los topic maps.
- Las restricciones necesarias relativas a su validación y corrección.

4.9.7 Tesauros

Etimológicamente viene del latín y su significación era tesoro. Es una definición que tiene poco que ver con el concepto que tiene hoy en la gestión documental, así la norma ISO 2788 / TC 46 dice que los tesauros se pueden definir según su función y según su estructura:

- Por su función, como un instrumento de control terminológico utilizado para trasponer a un lenguaje más estricto el idioma natural empleado en los documentos y por los indicadores.
- Por su estructura, es un vocabulario controlado y dinámico de términos que tienen entre ellos relaciones semánticas y genéricas y que se aplica a un dominio particular del conocimiento.

Los tesauros tienen como finalidad minimizar los inconvenientes que plantea el lenguaje natural en la recuperación de información. Al igual que los topic maps, se aplican como herramientas para la organización del vocabulario de un determinado ámbito científico. Contribuyen a dar coherencia a la indización, y su lenguaje es controlado en función de las necesidades y la materia.

El tesoro constituye una herramienta que facilita al usuario la comprensión del contexto de interpretación del concepto, en especial en la fase de búsqueda. El tesoro representa, por un lado, la estructura de los conocimientos transmitidos por los creadores de la información, y, por otro lado, la estructura cognitiva de los usuarios que formulan la expresión de búsqueda. En los tesauros, los descriptores, representan conceptos o nociones que pueden ser concretas, abstractas, entidades individuales o clases. Para expresar los conceptos, los tesauros contienen unidades léxicas llamadas descriptores y no descriptores.

Las relaciones que hay entre los términos de un tesoro son:

- Relaciones de equivalencia: es la relación entre descriptores o términos preferidos y los no descriptores o términos no usados en la indexación referentes a un mismo concepto. Se suelen considerar términos preferidos aquellos que más corrientemente son utilizados por los usuarios.
- Relaciones jerárquicas: es la relación vertical entre todos los descriptores de una misma clase, expresada en términos de subordinación de los conceptos. Estas relaciones se utilizan durante la búsqueda para enriquecer la formulación de la pregunta, añadiendo a la búsqueda uno o más descriptores superiores o inferiores para concretar la búsqueda.
- Relaciones asociativas: indican relación o uniones en la significación de los descriptores. Son relaciones simétricas entre dos descriptores, que son susceptibles de evocarse mutuamente por asociación de ideas. Las relaciones asociativas son las más difíciles de definir.

En cuanto a la elaboración de un tesoro, es una operación intelectual, en la cual la automatización sólo puede aportar una ayuda para las tareas de carácter más rutinario y repetitivo. Su elaboración supone el trabajo continuo de un grupo de especialistas en la materia durante un período de tiempo bastante amplio, que dependerá del volumen del tesoro, de los campos a cubrir y de la profundidad de la indexación. Para la compilación existen dos métodos:

- Inductivo: parte de cierto número de términos de los dominios considerados a través de enciclopedias, diccionarios técnicos, índices bibliográficos...
- Deductivo: cuyo vocabulario se extrae de los documentos, por indexación libre de cierto número de documentos, mediante el recuento de la frecuencia de aparición.

Ambos métodos se utilizan conjuntamente, existiendo siempre una normalización, aunque sea mínima con preferencia por la forma singular, masculina, forma sustantiva, escritura directa.

Obteniendo el vocabulario se procederá a su reducción por la eliminación de los términos no significativos y, el establecimiento de las relaciones entre los descriptores y entre estos y los no descriptores.

Una vez concluido el vocabulario se procederá a la evaluación del mismo. Con lo cual se procederá a su edición. Pero el tesoro no es algo definitivo y absoluto, este se debe ajustar a los progresos y a los propios usuarios y cambios en el sistema, por lo que cuando sea necesario debemos de recurrir a su mantenimiento, por medio de estadísticas de utilización de los descriptores.

En cuanto a la presentación del tesoro hay varias:

- Presentación alfabética: los descriptores y no descriptores se encuentran agrupados en una sola secuencia alfabética, acompañado de sus relaciones. No es la forma más habitual de encontrarlos.
- Presentación sistemática: consta de dos partes, una de ellas son las categorías o jerarquías y la otra un índice alfabético, que dirige a los usuarios a la sección semántica a la que el concepto pertenece. Considerándose esta última sección la parte principal del tesoro y el índice como auxiliar.
- Presentación gráfica: dispuestos como una figura que permite al usuario asociar los términos que se encuentran relacionados. Con dos tipos de presentación, los árboles y los flechados. También debe de constar de un índice alfabético, pues es difícil representar los no descriptores, notas de alcance. Dándose el caso de que la sección auxiliar del índice contiene más información que la parte en teoría principal, siendo esta gráfica de apoyo.

4.9.8 Web semántica

La web semántica (WS) es un área pujante en la confluencia de la Inteligencia Artificial y las tecnologías web que propone introducir descripciones explícitas sobre el significado de los recursos, para permitir que las propias máquinas tengan un nivel de comprensión de la web suficiente como para hacerse cargo de una parte, la más costosa, rutinaria, o físicamente inabarcable, del trabajo que actualmente realizan manualmente los usuarios que navegan e interactúan con la web.

La WS es una de las propuestas más completas en el ámbito de la gestión documental. La Web Semántica proporciona un marco general para la estructuración de ontologías y la vinculación de éstas entre sí. La unidad básica de las estructuras de tipo WS es la afirmación. La WS permite ensamblar e interrelacionar cantidades de estructuras de este tipo para crear una estructura más grande que codifique toda la información de un catálogo.

La Web Semántica es una Web extendida, dotada de mayor significado en la que cualquier usuario en Internet podrá encontrar respuestas a sus preguntas de forma más rápida y sencilla gracias a una información mejor definida. Al dotar a la Web de más significado y, por lo tanto, de más semántica, se pueden obtener soluciones a problemas habituales en la búsqueda de información gracias a la utilización de una infraestructura común, mediante la cual, es posible compartir, procesar y transferir información de forma sencilla. Esta Web extendida y basada en el significado, se apoya en lenguajes universales que resuelven los problemas ocasionados por una Web carente de semántica en la que, en ocasiones, el acceso a la información se convierte en una tarea difícil y frustrante.

La Web ha cambiado profundamente la forma en la que nos comunicamos, hacemos negocios y realizamos nuestro trabajo. La comunicación prácticamente con todo el mundo en cualquier momento y a bajo coste es posible hoy en día. Podemos realizar transacciones económicas a través de Internet. Tenemos acceso a millones de recursos, independientemente de nuestra situación geográfica e idioma. Todos estos factores han contribuido al éxito de la Web. Sin embargo, al mismo tiempo, estos factores que han propiciado el éxito de la Web, también han originado sus principales problemas: sobrecarga de información y heterogeneidad de fuentes de información con el consiguiente problema de interoperabilidad.

La Web Semántica ayuda a resolver estos dos importantes problemas permitiendo a los usuarios delegar tareas en software. Gracias a la semántica en la Web, el software es capaz de procesar su contenido, razonar con este, combinarlo y realizar deducciones lógicas para resolver problemas cotidianos automáticamente.

Berners-Lee (2001) afirma que la web semántica propone superar las limitaciones de la web actual mediante la introducción de descripciones explícitas del significado, la estructura interna y la estructura global de los contenidos y servicios disponibles en la WWW. Frente a la semántica implícita, el crecimiento caótico de recursos, y la ausencia de una organización clara de la web actual, la web semántica aboga por clasificar, dotar de estructura y anotar los recursos con semántica explícita procesable por máquinas. La figura 7 ilustra esta propuesta. Actualmente la web se asemeja a un grafo formado por nodos del mismo tipo, y arcos (hiperenlaces) igualmente indiferenciados. Por ejemplo, no se hace distinción entre la página personal de un profesor y el portal de una tienda on-line, como tampoco se distinguen explícitamente los enlaces a las asignaturas que imparte un profesor de los enlaces a sus publicaciones. Por el contrario en la web semántica cada nodo (recurso) tiene un tipo (profesor, tienda, pintor, libro), y los arcos representan relaciones explícitamente diferenciadas (pintor – obra, profesor – departamento, libro – editorial).

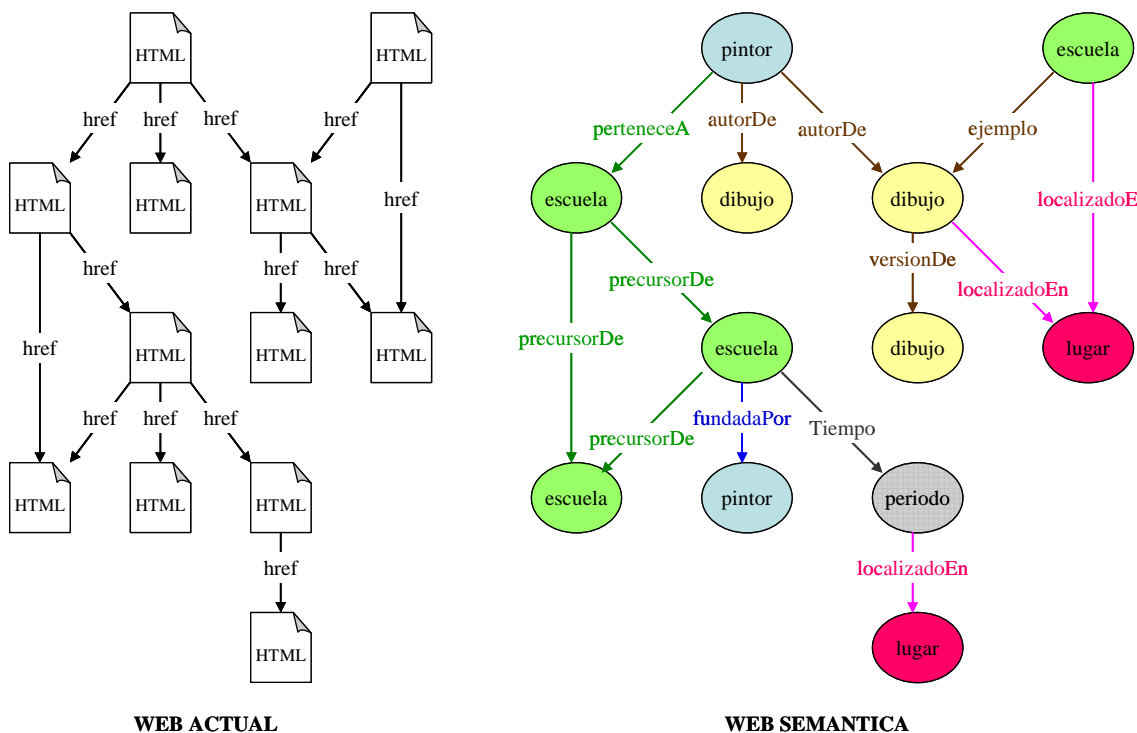


Figura 7: web actual y web semántica

La web semántica mantiene los principios que han hecho un éxito de la web actual, como son los principios de descentralización, compartición, compatibilidad, máxima facilidad de acceso y contribución, o la apertura al crecimiento y uso no previstos de antemano. En este contexto un problema clave es alcanzar un entendimiento entre las partes que han de intervenir en la construcción y explotación de la web: usuarios, desarrolladores y programas de muy diverso perfil.

Por último, la web no solamente proporciona acceso a contenidos sino que también ofrece interacción y servicios (comprar un libro, reservar una plaza en un vuelo, hacer una transferencia bancaria, simular una hipoteca). Los servicios web semánticos son una línea importante de la web semántica, que propone describir no sólo información sino definir ontologías de funcionalidad y procedimientos para describir servicios web: sus entradas y salidas, las condiciones necesarias para que se puedan ejecutar, los efectos que producen, o los pasos a seguir cuando se trata de un servicio compuesto. Estas descripciones procesables por máquinas permitirían automatizar el descubrimiento, la composición, y la ejecución de servicios, así como la comunicación entre unos y otros.

4.9.9 Tendencias futuras

Existen más normas que las enumeradas en los apartados anteriores, como RDF y OWL e iniciativas (por ejemplo el ABC Ontology and Model) relacionadas.

Es importante reconocer que muchos de estos avances retoman ideas provenientes de la informática, y especialmente la rama de esta disciplina conocida como la "inteligencia artificial". Davis (1993) expresa que esta área de estudio aun está muy lejos de lograr su propósito de crear máquinas pensantes; sin embargo, como parte de esta búsqueda, ha

avanzado sustancialmente en la "representación del conocimiento", es decir, en la elaboración de estructuras de información como las que se ha mencionado.

El anuncio de estos cambios no debe resultar desconcertante dentro del mundo de la organización documental. Este cambio de paradigma no debe implicar la pérdida de logros ya consolidados en la sistematización y normalización en la codificación del conocimiento, sino la construcción de herramientas de consulta que permitirán ir más allá del universo de posibilidades hoy conocido. Asimismo, para aprender nuevos sistemas de gestión documental que a primera vista parecen bastante complicados, tan solo es necesario un poco de práctica para poder emplear las nuevas metodologías con facilidad, y su manejo será aun más sencillo una vez que estén disponibles las herramientas informáticas "amigables" que se pueden esperar que surjan.

El nuevo panorama que se abre trae a la vez problemas y soluciones que vuelven el trabajo con documentos un quehacer cada vez menos burocrático y más interesante, y evidencian su carácter de labor de investigación.

Sin duda este nuevo paradigma podrá ser acompañado de cambios en las configuraciones sociales del trabajo de gestión documental en específico y de la investigación en general.

4.10 El lenguaje de marcado

Un lenguaje de marcado es un conjunto de reglas que establecen qué tipo de marcas han de ser utilizadas, de qué modo se distinguen las marcas del texto del documentos y cómo se insertan éstas (la gramática y su sintaxis), y cuáles son las marcas permitidas en cada una de las partes del texto. De forma genérica, Martín y Rodríguez (2000) distinguen dos tipos de lenguajes de marcado:

- Los lenguajes de marcado procedimentales: orientados a la presentación de los documentos, especifican cómo debe ser procesado el texto para su salida a través de diversos medios (pantalla ordenador, impresora, etc.) Estos lenguajes no aportan información de tipo semántico o estructura; son poco flexibles, dado que cualquier cambio en la presentación del documento implica modificar su marcado; y suelen ser lenguajes específicos de un sistema de procesamiento propietario, lo cual reduce la "portabilidad" de dichos documentos. Algunos ejemplos de estos lenguajes son RTF (Rich Text Format) de Microsoft y PDF (Portable Document Format) de Adobe.
- Lenguajes de marcado descriptivos: orientados a la descripción formal y de contenido de los documentos. Estos lenguajes aportan información sobre la estructura del documento y describen el contenido de información del mismo, además, son lenguajes más flexibles, que diferencian entre el contenido real del documento y su representación. Algunos ejemplos son SGML, HTML, XML, etc.

4.10.1 El lenguaje XML

XML es un proyecto del World Wide Web Consortium (W3C) y su desarrollo está coordinado por el XML working group. Se trata de un subconjunto (o simplificación) adaptado de SGML que tiene la intención de aprovechar sus máximas ventajas posibles, pero permitiendo su implementación en Internet.

El lenguaje XML consta de cuatro especificaciones (recomendaciones de W3C):

- DTD (“Document Type Definition”): Definición de tipo de documento. Se trata de un archivo que encierra una definición formal de un tipo de documento y, a la vez, especifica la estructura lógica de cada documento. El DTD en XML es opcional. En tareas sencillas no es necesario construir una DTD, entonces se trataría de un documento “bien formado” (well-formed). Si se construye una DTD, entonces será un documento “validado” (valid).
- XSL (“eXtensible Stylesheet Language”): Lenguaje de estilo para XML. Se trata de un lenguaje para elaborar hojas de estilo. Consta de tres partes:
 - “XSL Transformations” (XSLT), es un lenguaje para transformar documentos XML;
 - “XML Path Language” (XPath), es un lenguaje de expresión usado por XSLT para acceder o referir partes de un documento XML (XPath se usa también en la especificación XML Linking).;
 - “XSL Formatting Objects”, es un vocabulario para especificar la semántica del formato.
- XLL (“eXtensible Linking Language”): Lenguaje de enlaces en XML. Define el modo de actuación entre diferentes enlaces. Se considera un subconjunto de HyTime (ISO 10744) y sigue algunas especificaciones de TEI. XLL tiene dos importantes componentes: Xlink y Xpointer, con los que se va mucho más allá de los enlaces simples soportados por HTML, los enlaces pueden ser bidireccionales, múltiples (anillos, múltiples ventanas, etc.), agrupados (múltiples orígenes).
- XUA (XML User Agent): Estandarización de navegadores XML. Se aplica a los navegadores para que sean capaces de reconocer todas las especificaciones XML.

Fernando Santamaría (1998) destaca algunas de las características esenciales de XML y su influencia directa con el desarrollo de servicios de información vía web.

- XML es una arquitectura abierta y extensible. No necesita versiones para funcionar en futuros navegadores.
- XML posee mayor consistencia, homogeneidad y amplitud de los identificadores descriptivos del documento (RDF), en comparación con los atributos de la etiqueta <META> de HTML.
- XML permite agrupar una amplia variedad de datos y aplicaciones, desde páginas web hasta bases de datos.
- XML consigue que la estructura de la información resulte más accesible, por tanto, los motores de búsqueda devolverán respuestas más adecuadas y precisas.
- XML permite el desarrollo de búsquedas personalizables para robots y agentes inteligentes.
- XML desarrolla ampliamente el concepto de “hipertexto” mediante enlaces bidireccionales, enlaces que pueden especificarse y gestionarse desde fuera del documento, hiperenlaces múltiples, enlaces agrupados, atributos para enlaces, etc.
- XML facilita la exportabilidad a otros formatos de publicación (papel, web, cd-rom, etc.)

4.10.2 Los lenguajes de marcado y la codificación del conocimiento

Cada lenguaje de marcado de información utiliza mecanismos diferentes para la representación del conocimiento contenido en un documento. En este apartado vamos a tratar de los dos lenguajes de marcado más extendidos para la publicación de contenidos en Internet: HTML y XML. Para ambos lenguajes se determinan los mecanismos que podríamos denominar de representación del conocimiento documental o codificación de la meta información documental:

- “Dublin Core” (DC) en HTML. DC es el formato de metainformación más divulgado y más citado a nivel general, al menos en el ámbito estrictamente de gestión documental. Desde sus comienzos (1995), el DC fue diseñado para promover un estándar de propósito general, sencillo y descriptivo de los recursos web de cualquier materia.
- “Resource Description Framework” (RDF) en XML. RDF es sin duda alguna el estándar en desarrollo más importante para la descripción de contenidos web. Su importancia reside, por un lado en la entidad Consorcio Web (W3C) quien fomenta el proyecto, y por otro, el hecho de ser una aplicación de metadatos que utiliza XML a fin de proporcionar un marco estándar para la interoperabilidad

entre distintos modelos de metadatos para la representación de los conocimientos (metainformación) contenido en un documento (como por ejemplo el DC, que admite las especificaciones RDF a partir de la versión DC5). Desde que RDF se convirtió en una recomendación de W3C en febrero de 1999, se han desarrollado un buen número de herramientas que permiten trabajar con RDF de una forma más eficiente. RDF ofrece una variedad de aplicaciones, tales como:

- catálogos de biblioteca y directorios web (ejemplos: Dublin Core Metadata Initiative, OCLC Connexion, Open Directory Project, etc.)
- categorización y gestión de listas de noticias, software y contenidos (PICS, XMLNew, UK Mirror Service)
- colecciones de música, fotos y eventos (MusicBrains metadata initiative, RDFPic, etc.)

4.10.3 Desarrollo de servicios de información basados en XML

Las organizaciones que prestan servicios de información deben diseñar sistemas eficaces para gestionar y difundir la información. Con la aplicación de las tecnologías de la información en el desarrollo e implementación de sistemas de información y la utilización de la tecnología web, un porcentaje muy alto de los servicios de información se fundamentan en información electrónica. En este contexto la familia de normas sgml resulta clave, puesto que: asegura la consistencia (en la producción, el procesamiento, el almacenamiento y la distribución) y proporciona una enorme flexibilidad (en la presentación y en el formato, por ejemplo).

- XML en los servicios de gestión documental. Los profesionales de la información están muy interesados en el potencial de XML para la organización de la información. Uno de los campos de mayor actividad consiste en la sustitución de MARC usando tecnología XML, de hecho la Library of Congress y otras organizaciones ya han llevado a cabo este tipo de proyectos.
- XML en servicios de información comercial. Cada día son más numerosos los proyectos xml que se están emprendiendo en el ámbito de los servicios de información comercial. Se trata de aplicaciones relacionadas con dominios muy variados: banca, telecomunicaciones, revistas electrónicas, etc. (por ejemplo el proyecto OASIS: Organization for the Advancement of Structural Information Organization <<http://www.oasis-open.org>>). Esta proliferación de aplicaciones xml no significa que éste vaya a sustituir a sgml. En principio, mantener una o varias bases de datos sgml y filtrarlas a xml para las operaciones de indización, otros tratamientos de los datos y distribución en la web, se conforma como una estrategia a largo plazo con gran posibilidad de éxito, según Peis y Moya (2000). De hecho, la mayoría de las bibliotecas digitales mantienen sus datos en sgml, pero distribuyen la información en xml.

4.10.4 XML en la gestión del conocimiento organizacional

La literatura científica reciente se centra en las ventajas de XML para el desarrollo de sitios web y su aplicación a las tecnologías de búsqueda y recuperación de información web. En esta ocasión es necesario tener en cuenta los usos de XML para la gestión del conocimiento organizacional. XML es un lenguaje que presenta simultáneamente el contenido para ser publicado en Internet y describe este contenido de forma que otro software puede entender y usar este conocimiento. La clave de XML es que, en oposición a HTML, proporciona información sobre el significado de los datos, lo que va a facilitar el procesamiento automático de la información contenida en la información (meta información) y la consiguiente gestión del conocimiento.

XML se empieza a considerar como “la próxima revolución en la gestión del conocimiento”, y las organizaciones están empezando a entender el potencial de esta tecnología para el desarrollo de arquitecturas de información corporativas.

La tecnología XML por sí misma no reportan ningún valor a la organización, su valor depende de cómo se use esta tecnología dentro de la organización. Su implementación no debe ser departamental e, idealmente, debería incluir a partners estratégicos y otras organizaciones con las que se necesita compartir datos e información. Un ejemplo claro de los beneficios de XML en la gestión de la información corporativa consiste en la creación de portales corporativos que ofrecen prestaciones para diferentes funciones de negocio, tales como: ERP, datwarehousing, sistemas de apoyo a la decisión y gestión del conocimiento organizacional.

Entre las aplicaciones estratégicas que XML puede tener dentro de una organización podemos destacar:

- Cadena de suministro: La gestión eficiente del conocimiento organizacional depende del acceso a la información externa. XML puede ayudar a mejorar el funcionamiento de la cadena de suministro y la extranet de la organización. Muchas organizaciones necesitan compartir información y no siempre tienen una vía para hacerlo, puesto que tienen infraestructuras tecnológicas diferentes y no compatibles con algunos agentes de la cadena de suministro o con los clientes. XML puede proporcionar una forma estándar para intercambiar información con proveedores, distribuidores y clientes, lo cual se convierte en una diferencia competitiva para la organización capaz de gestionar la cadena de suministro mediante XML.
- Retorno de Inversiones (ROI): Los documentos XML pueden producir diferentes outputs y una vez que la información organizacional está organizada en documentos XML, se pueden elaborar multitud de productos y servicios de información vía web, así como en diferentes plataformas: PDA, telefonía móvil, televisión digital, etc. La creación de un corpus de conocimiento organizacional en XML supone un gran esfuerzo de inversión, pero pronto reporta un ahorro en los costes directos relacionados con la inversión en tecnologías de la información, la formación de los empleados y los servicios de soporte.

- Gestión integral del conocimiento: XML posee un conjunto de características que lo convierten en un aliado para la gestión del conocimiento organizacional. Si se combina el potencial de XML y las posibilidades de RDF, como herramienta para la gestión del conocimiento que permite organizar, interrelacionar, clasificar y anotar el conocimiento contenido en los documentos XML, se puede obtener un incremento considerable del valor añadido de los datos almacenados en los documentos XML y conlleva una gestión automática del conocimiento explícito de la organización.