Análisis de características del ambiente creativo en empresas de Manizales con técnicas KDD

Germán Augusto Osorio Zuluaga

Universidad Nacional de Colombia – Sede Manizales
Facultad de Ingeniería y Arquitectura
Departamento de Ingeniería Eléctrica, Electrónica y Computación
Manizales
2009

Análisis de características del ambiente creativo en empresas de Manizales con técnicas KDD

Germán Augusto Osorio Zuluaga

Trabajo de grado para optar al título de
Magíster en Ingeniería — Automatización Industrial

Director
Profesor Néstor Darío Duque Méndez PhD

Universidad Nacional de Colombia – Sede Manizales
Facultad de Ingeniería y Arquitectura
Departamento de Ingeniería Eléctrica, Electrónica y Computación
Manizales
2009

Tabla de Contenido

Tabla de Contenido 3
Lista de figuras 5
Lista de tablas 6
Agradecimientos 7
Dedicatoria 8
Resumen 9
Abstract 10
Parte I: Preliminares 11
1. Introducción 12
2. Objetivos 14
2.1 Objetivo General 14
2.2 Objetivos Específicos 14
Parte II: Marco Teórico 15
3. Creatividad 16
3.1 Historia del concepto 16
3.2 Definición de Creatividad 18
3.3 Característica medioambiental de la creatividad 21
3.4 Conclusión del capítulo 23
4. Descubrimiento de conocimiento en bases de datos 24
4.1 Definición de KDD 24
4.2 ¿Qué es la minería de datos? 25
4.3 El proceso de extracción de conocimiento 25
4.4 Tareas de minería de datos 29
4.5 Técnicas de minería de datos 29
4.5.1 Fiabilidad de los datos: Alfa de Cronbach 30
4.5.2 Representación gráfica: diagramas de dispersión 31
4.5.3 Detección de datos atípicos 32
4.5.4 Reducción de dimensión: Análisis de Componentes Principales 32
4.5.5 Reducción de dimensión: Escalamiento multidimensional 33
4.5.6 Análisis de Conglomerados 34
4.6 Conclusión del capítulo 36

5. Estado del arte 37
5.1 Revisión del estado del arte 37
5.2 Conclusión del capítulo 39
Parte III: Marco Experimental 41
6. Resultados de las pruebas 42
6.1 Aplicación de la metodología 42
6.1.1 Desarrollar un entendimiento del dominio de la aplicación 42
6.1.2 Seleccionar y crear un conjunto de datos en el cual se ejecutará el descubrimiento. 43
6.1.3 Limpieza y preprocesamiento de los datos. 44
6.1.4 Elección de tareas y algoritmos de minería de datos y su empleo 48
6.1.4.1 Experimento 1 48
6.1.4.2 Experimento 2 50
6.1.4.3 Experimento 3 55
6.1.4.4 Experimento 4 57
6.1.4.5 Experimento5 62
6.1.4.6 Experimento 6 65
6.1.4.7 Experimento 7 67
Parte IV: Discusión y Conclusiones 70
7. Discusión 71
7.1 Discusión sobre los resultados del experimento 1 71
7.2 Discusión sobre los resultados del experimento 2 71
7.3 Discusión sobre los resultados del experimento 3 72
7.4 Discusión sobre los resultados del experimento 4 73
7.5 Discusión sobre los resultados del experimento 5 74
7.6 Discusión sobre los resultados del experimento 6 75
7.7 Discusión sobre los resultados del experimento 7 75
8. Conclusiones 76
Bibliografía 78

Lista de figuras

Ilustración 1: Mapa conceptual de la evolución histórica del concepto de creatividad 19
Ilustración 2: Mapa conceptual de la definición de creatividad según Huidobro, 2002 20
Ilustración 3: Mapa conceptual de modelo medioambiental de la creatividad según González, 2006 22
Ilustración 4: Proceso general de KDD (Bramer, 2007) 24
Ilustración 5: Proceso KDD 26
Ilustración 6: Diagrama de dispersión 31
Ilustración 7: Dendograma 36
Ilustración 8: Diagrama sobre las 15 características 45
Ilustración 9: Diagrama de dispersión sobre las 7 primeras características 45
Ilustración 10: Proyección de los datos sobre las primeras componentes 46
Ilustración 11: Coeficientes de los vectores propios sobre tres componentes principales .. 47
Ilustración 12: Nubes de datos proyectadas en dos empresas 47
Ilustración 13: Agrupamiento jerárquico de los datos 48
Ilustración 14: Proyección de datos sobre las 2 primeras componentes, 6 grupos calculados con K-medias 51
Ilustración 15: Escalamiento multidimensional de todos los datos 52
Ilustración 16: Observaciones atípicas proyectadas sobre 2 componentes 52
Ilustración 17: Agrupamiento jerárquico sin atípicos 53
Ilustración 18: Proyección de datos sobre las 2 primeras componentes, 3 grupos calculados con K-medias 56
Ilustración 19: Agrupamiento jerárquico sin atípicos todas las 52 variables 57
Ilustración 20: Gráfico de sedimentación sobre los autovalores calculados con PCA 61
Ilustración 21: Diagrama de dispersión sobre la proyección de los datos sobre los 10 primeros componentes 61
Ilustración 22: Proyección de los datos (todas las variables) sobre los dos primeros componentes 62
Ilustración 23: Proyección de lo datos sobre dos componentes por sectores 64
Ilustración 24: Proyección de datos sobre las dos primeras componentes por empresas del sector Servicios 66
Ilustración 25: Proyección de datos sobre las dos primeras componentes por empresas del sector Insumos 66
Ilustración 26: Proyección de datos sobre las dos primeras componentes por empresas del sector Alimentos y Bebidas 67
Ilustración 27: Proyección de datos sobre las dos primeras componentes por empresas del sector Productos de consumo 67
Ilustración 28: Proyección de los datos sobre las dos primeras componentes por nivel jerárquico 68

Lista de tablas

Tabla 1: Algunas técnicas de Minería de Datos 30
Tabla 2: Alfa de Cronbach calculado para los datos de la encuesta 44
Tabla 3: Matriz de coeficientes de los siete primeros componentes 46
Tabla 4: Suma de cuadrados dentro de los grupos 49
Tabla 5: Estadísticos de centroides para 6 grupos generados por K medias en 10 ejecuciones 50
Tabla 6: Suma de cuadrados dentro de los grupos, sin atípicos 53
Tabla 7: Estadísticos de centroides para 8 grupos generados por K medias en 10 ejecuciones 54
Tabla 8: Estadísticos de centroides para 5 grupos generados por K medias en 10 ejecuciones 54
Tabla 9: Estadísticos de centroides para 3 grupos generados por K medias en 10 ejecuciones 55
Tabla 10: Numero de grupos sin atípicos todas las variables (calculados por K medias) .. 58
Tabla 11: Estadísticos de centroides para 5 grupos generados por K medias en 10 ejecuciones, consideradas todas las 52 variables 59
Tabla 12: Estadísticos de centroides para 3 grupos generados por K medias en 10 ejecuciones, consideradas todas las 52 variables 59
Tabla 13: Estadísticos de centroides para 2 grupos generados por K medias en 10 ejecuciones, consideradas todas las 52 variables 60
Tabla 14: Aporte individual de cada respuesta a la medición del ambiente creativo 63
Tabla 15: Varianza explicada por cada componente por sector 64
Tabla 16: Peso de la primera componente sobre las dimensiones ambientales y total por sector 65
Tabla 17: Varianza explicada por cada componente por nivel jerárquico 68
Tabla 18: Peso de la primera componente en las dimensiones ambientales y el total por nivel jerárquico 69

Agradecimientos

¡A Jesucristo, Señor y Dios nuestro y a María Santísima, nuestra tierna madre, por todo!

- A mis padres y mis hermanos, por sentar las bases de mi ser y por su apoyo permanente e incondicional

- A la Universidad Nacional de Colombia, institución que me ha permitido desarrollarme personal, académica, profesional y laboralmente.

- Al profesor Néstor Darío Duque Méndez, por su orientación, apoyo y ánimo en el desarrollo de esta tesis.

- A Luís Gonzalo Sánchez Giraldo, por su apoyo y asesoría técnica, al igual que por su acogida y amistad.

- Al profesor Mauricio Orozco, por su asesoría y disposición a apoyar mi trabajo; lo mismo que por su buen compañerismo.

Dedicatoria

¡A Piedad Eugenia, mi tierna y amorosa compañera de camino!

“Initium sapientiæ timor Domini”
Prov 1,7

"Vocatus atque non vocatus, Deus aderit”
Antiguo proverbio espartano, popularizado por Carl Gustav Jung

Resumen

En este trabajo se presentan los resultados de la aplicación de técnicas de Minería de Datos, (DM, por sus siglas en inglés), enmarcadas en el proceso de Descubrimiento de Conocimiento en Bases de Datos (KDD), sobre la información de una investigación realizada por la Universidad Nacional de Colombia - Sede Manizales en los años 2004 y 2005 en veintitrés empresas de la ciudad, con el fin de determinar el estado de los ambientes creativos para la innovación, desde las dimensiones psicosocial, didáctica y física.

Se consideran las técnicas de Análisis de Componentes Principales (PCA), Escalamiento Multidimensional, agrupamiento por K-medias y agrupamiento jerárquico.

En las pruebas realizadas, se determinó la ponderación de las variables latentes de las dimensiones didáctica, física y psicosocial que varían un poco respecto de la dada de manera empírica por los investigadores. Adicionalmente se logró determinar que características tienen mayor influencia en la determinación de un adecuado ambiente que favorezca la creatividad, como son la comunicación, la formación, el aspecto funcional, la característica técnica, el reconocimiento, la motivación-satisfacción y la apertura.

Igualmente, se pudo determinar que las dimensiones didáctica, física y psicosocial influyen de manera distinta en los sectores y niveles organizacionales.

Por otro lado, se demuestra al menos en este caso, la suposición que se tenía sobre la posibilidad de descubrir nuevo conocimiento con el uso de estas técnicas aplicadas en datos de investigaciones anteriores que por razones diversas no las usaron.

Finalmente, se muestra las grandes posibilidades de la aplicación de la ingeniería en ciencias sociales y del comportamiento.

Abstract

This document presents the results of the application of Data Mining (DM) techniques framed in the process of Knowledge Discovery in Databases (KDD), information on research conducted by Universidad Nacional de Colombia - Sede Manizales in 2004 and 2005 in twenty-three companies in the city, in order to determine the status of creative environments for innovation, from the psychosocial, didactic and physical dimensions.

Techniques considered are principal component analysis (PCA), multidimensional scaling, clustering by K-means and hierarchical clustering.

In tests, it was determined by weighing that the latent variables of the didactic, physical and psychosocial dimensions, vary slightly on the empirically given by the researchers.

Additionally it was determined that characteristics have greater influence in establishing an appropriate environment conducive to creativity, such as communication, training, the functional aspect, the technical feature, recognition, motivation, satisfaction and openness.

Similarly, it was determined that the didactic, physical and psychosocial dimensions, impact differently on areas and organizational levels.

On the other hand, it is shown at least in this case, our assumption about the possibility of discovering new knowledge through the use of these techniques applied in previous research data which, for various reasons, did not use them.

Finally, it is shown the great possibilities of implementing engineering in social and behavioral sciences.

Parte I: Preliminares

1.Introducción

El descubrimiento de conocimiento en bases de datos (KDD, por sus siglas en inglés) en los últimos años ha ganado preponderancia, se viene desarrollando y utilizando ampliamente ya como una disciplina con un cuerpo teórico muy estructurado. Uno de sus componentes más importantes es la Minería de Datos que integra técnicas de análisis de datos y extracción de modelos.

La Minería de Datos se basa en varias disciplinas, algunas de ellas tan antiguas como la estadística. Por ello, lo novedoso no son tanto sus técnicas sino su fin, el cual es el de extraer conocimiento.

Se viene utilizando considerablemente en el sector financiero y banca, análisis de mercados, seguros, procesos industriales, medicina, biología, telecomunicaciones, ingeniería, investigación en ciencias duras, entre otras.

También se ha aplicado en investigación en ciencias sociales y del comportamiento, pero en menor medida. La creatividad, como uno de sus objetos de estudio, no es ajena a esta situación: a nivel mundial son escasas las investigaciones en las que se haga uso de la metodología y técnicas de KDD en la evaluación de la creatividad.

Y qué decir a nivel nacional y regional. En el año 2005, el GTA PROCREA adscrito a la Facultad de Ingeniería de la Universidad Nacional – Sede Manizales, desarrolló una investigación en 23 empresas consideradas estratégicas en cuatro sectores productivos, referente a valoración del ambiente creativo en cada una de ellas. Se recolectó gran cantidad de información aportada por una encuesta aplicada a los empleados en los niveles estratégico, táctico y operativo y para su cuantificación se usaron estadísticos básicos univariados, sin profundizar en ellos.

El estudio de la creatividad y la innovación toman gran importancia, ya que en un ambiente tan competitivo a nivel global y local, las empresas deben ser muy creativas e innovadoras para lograr sobrevivir y crecer. Así lo señala Michael Porter, autor de "La ventaja competitiva de las naciones " cuando afirma que "la competitividad de una nación depende de la capacidad de su industria para innovar y mejorar. Las empresas consiguen ventajas competitivas si consiguen innovar".

Este trabajo aborda, haciendo uso de la metodología de KDD, el problema de extraer conocimiento adicional sobre los datos de la investigación mencionada, con la suposición previa que las técnicas utilizadas anteriormente no permitían obtener todo el conocimiento oculto de los datos.

El documento está organizado como se menciona a continuación:

En el capítulo 3, se da una contextualización sobre creatividad, especialmente la evolución histórica de su concepto, la definición propiamente dicha y por último, se profundiza un poco sobre el aspecto medioambiental de la creatividad, el cual, es el objeto de estudio en esta tesis.

Luego, en el capítulo 4, se aborda una sustentación teórica general sobre KDD. Igualmente se presentan algunas de técnicas de minería de datos, usadas en el marco experimental de esta tesis.

Posteriormente en el capítulo 5, se hacer una revisión del estado del arte sobre el uso de herramientas KDD para evaluar la creatividad. En el capítulo 6, se presentan y describen las pruebas realizadas para la extracción de conocimiento. A continuación en los capítulos 7 y 8, se plantean la discusión y las conclusiones respectivamente del trabajo realizado.

2. Objetivos

2.1 Objetivo General

Analizar e identificar, con técnicas KDD, las características que más influyen en la determinación del ambiente creativo y la innovación para la mejora de la productividad en las empresas de Manizales y determinar patrones estructurales existentes.

2.2 Objetivos Específicos

1. Evaluar diferentes técnicas KDD descriptivas y predictivas e implementar las más adecuadas para el análisis y búsqueda de patrones estructurales.

2. Identificar las variables más determinantes en la generación de ambientes creativos en las empresas de Manizales.

3. Determinar patrones estructurales existentes por nivel laboral, empresa y/o sector económico.

4. Validar los modelos obtenidos.

Parte II: Marco Teórico

3. Creatividad

En este capítulo se presenta una contextualización sobre creatividad.

Se parte de la evolución histórica del concepto de creatividad. A continuación se analiza el concepto propiamente dicho. Y, finalmente, se abordan los aspectos medioambientales que influyen en la creatividad, que es la temática específica de este trabajo.

3.1 Historia del concepto

En [1] se hace un recuento histórico del concepto de creatividad vinculada a la evolución del arte, partiendo de Tatarkiewicz [2], la cual se complementa con [3] que también referencia dicho autor:

Para el mundo griego de la antigüedad, las artes interpretaban el universo, máxima expresión de la perfección, en donde lo creado venía de los dioses y por consiguiente, para ellos nada surgía de la nada. El artista se consideraba como un medio para producir su obra a través de la imitación (mimesis). No era necesario usar conceptos como "crear" o "creador" porque "fabricar" les bastaba. No se valoraba la libertad expresiva del artista para dilucidar las normas y los modelos perpetuos que regían la perfección de la naturaleza, además de su habilidad en aprovecharlas en los distintos ámbitos del quehacer humano.

Esto se mantuvo durante el periodo romano, sólo que se flexibiliza el ideal de la búsqueda de la verdad mediante uso más amplio de las ideas de imaginación e inspiración fuera del ámbito de la poesía. No obstante, estas ideas se vinculan a un don dados por los dioses a algunos hombres, en muchos casos manifestados por medio de la locura, permitiéndoles acceder a la perfección y al conocimiento de la verdad de una manera más espontánea y libre.

En la Antigüedad, se negó la creatividad, planteando la "emanación" como motor de la creación: nuevamente se señala que nada puede surgir de la nada. Más adelante, esta visión se reemplaza por el filósofo Lucrecia que indica que la creatividad existe, pero el hombre es incapaz de ella.

En el Medioevo, el concepto de creación de los padres de la Iglesia y de los escolásticos era el mismo que en el filósofo Lucrecia. Hay un cambio conceptual importante, pues la idea de crear algo que previamente no existiese se incorpora a la cultura occidental. Sin embargo, el concepto y el término latino se asocia exclusivamente a la capacidad divina de crear algo de la nada, sin poderse aplicar a los campos de la actividad humana, la cual está limitada a la observación de la obra divina con el fin de imitarla.

Tanto en la Antigüedad como en el Medioevo, la creatividad como concepto bajo la perspectiva histórica era "interpretada como divina".

En el periodo del Renacimiento, a pesar de su gran producción artística, no se definió conscientemente el concepto. En este periodo, el deseo de encontrar expresión a la naciente conciencia de independencia y libertad se manifestó a través del descubrimiento de su capacidad de producir cosas nuevas y de la necesidad de hacerlo. Aunque surge la noción de la novedad como valor y capacidad del ser humano, aún no se utiliza la palabra "creador", el concepto de creatividad, empieza a gestarse semánticamente.

Ya en el siglo XVII, Marciej Kazimierz Sarbiewski plantea que el poeta inventa, construye según su estilo, crea algo nuevo, el poeta crea tal como lo hace Dios. De modo similar, escribe Baltasar Gracián: "Es el arte complemento de la naturaleza y otro segundo creador que la embellece y aún pretende superarla con sus obras. Se precia el artista de haber añadido otro mundo artificial al primero, de suerte que al ser el artificio gala de lo natural y realce de su llaneza, obra siempre milagros" [4]. Si bien por primera vez se emplea el término "creador" para referirse al hombre; el primer autor lo vincula exclusivamente al poeta y el segundo lo extiende sólo al campo del arte.

No obstante lo anterior, fue en este siglo cuando existió mayor resistencia a la agregación del término creatividad como fruto de la mente humana: según Batteux, "la mente humana, estrictamente hablando, no puede crear; todos sus productos llevan el estigma de su modelo". Esta resistencia se debía a tres aspectos de origen: 1. "Un origen era lingüístico: la expresión "creación" estaba reservada en el uso contemporáneo a la creación ex nihilo (de la nada) que era inaccesible al hombre. 2. La segunda fuente era filosófica: la creación era un acto misterioso, y la psicología de la Ilustración no admitía misterios. 3. Finalmente el tercer origen era artístico: los artistas de la época estaban sujetos a reglas, y la creatividad parecía irreconciliable con las reglas".

Ya en el siglo XVIII, el concepto de creatividad fue apareciendo más mayor frecuentemente en la teoría del arte, unido al concepto de imaginación.

En el siglo XIX, se vencieron todas las resistencias anteriores y es cuando definitivamente se habla del hombre como creador y se usan acepciones como el adjetivo "creativo" y el sustantivo creatividad, considerantes antes como superfluas. Y no sólo se reconoció la creatividad, sino que el concepto "creador" fue considerado de forma unitaria para el artista y poeta.

El Romanticismo marca este punto de inflexión, ya que se trata de una reacción en diversos ámbitos: estéticos, políticos, culturales, contra el excesivo racionalismo y utilitarismo de la Ilustración. Se construye un concepto de creatividad en el que por primera vez se excluye el requisito "a partir de la nada". La creatividad se define a partir de la capacidad de producir cosas nuevas; concepto vinculado exclusivamente al arte y a los artistas.

A principios del siglo XX, se presenta otro momento clave cuando se usó el término en ciencias, cuando era propio del arte.

Por último, a través de la evolución del concepto en el tiempo ha ido ampliándose su uso en diversos ámbitos de las producciones humanas, guardando distingos a la hora de hacer referencia a la pertinencia de la acepción en los campos del arte, la ciencia y la técnica.

El mapa conceptual de la *Ilustración 1*, recoge de manera sintética, la evolución del concepto de creatividad.

3.2 Definición de Creatividad

De acuerdo con [5], desde los inicios del estudio de la creatividad a principios del siglo XX, se fueron configurando unos referentes que son los que han ido conformando la creatividad como un fenómeno peculiar, diferenciado de otras capacidades humanas. Estos referentes son *la persona, el proceso, el producto* y *el ambiente o medio*. Con base en esos aspectos, definidos en primer lugar por Mooney en 1963 y MacKinnon en1970, se presentan las definiciones planteadas por [6], las cuales son sintetizadas en la *Ilustración 2*.

Desde *el punto de vista de la persona*, se define la creatividad como "una serie de rasgos de personalidad que se pueden clasificar en:

- Intelectuales: caracterizados por el equilibrio entre la apertura y el cierre, la razón y la imaginación, la fluidez y la concentración.
- No intelectuales, de personalidad, entre los que destacan el individualismo, la originalidad y el anticonvencionalismo.
- No intelectuales, motivacionales, fundamentalmente la motivación intrínseca (o por la tarea) y los intereses exploratorios y estéticos.

Desde *el punto de vista exclusivamente del proceso*: Creatividad es el "proceso cognitivo que comienza con la necesidad de interpretar una situación que supone un fallo o hueco en el funcionamiento o conocimiento acerca de algo, para lo cual se inicia la generación de soluciones, que se van comparando sucesivamente con una meta, y se continua hasta ejecutar una serie de medidas que logran dar respuesta al fallo o rellenar el hueco". La mente que es capaz de llevar a cabo este proceso responde a un modelo tal que aborda sus tareas teniendo en cuenta la dualidad, la regulación y la adaptabilidad y presenta una organización sistémica y un modo de funcionamiento flexible y reflexivo.

Desde el punto de vista del producto se define la creatividad como "la producción de algo nuevo y adecuado, que solucione un problema que era vago o estaba mal definido, suponga rareza estadística y una transformación radical de un estado anterior. Además, dicho producto debe tener suficiente trascendencia y producir un impacto". Evidentemente, dependiendo del nivel de creatividad exigido o el campo de aplicación (las artes o las ciencias), diferirán los criterios necesarios para considerar creativo a un producto.

Desde el punto de vista del contexto se define la creatividad como: "Una serie de circunstancias que rodean a la persona y al producto creativos y que se caracterizan por la disponibilidad de recursos económicos, formativos y culturales, por la presencia de modelos o parangones a los que imitar, y por un entorno familiar y social carente de obstáculos, dado que favorece y reconoce las conductas individualistas, innovadoras y creativas y que no ejerce una presión excesiva por obtener logros prematuros".

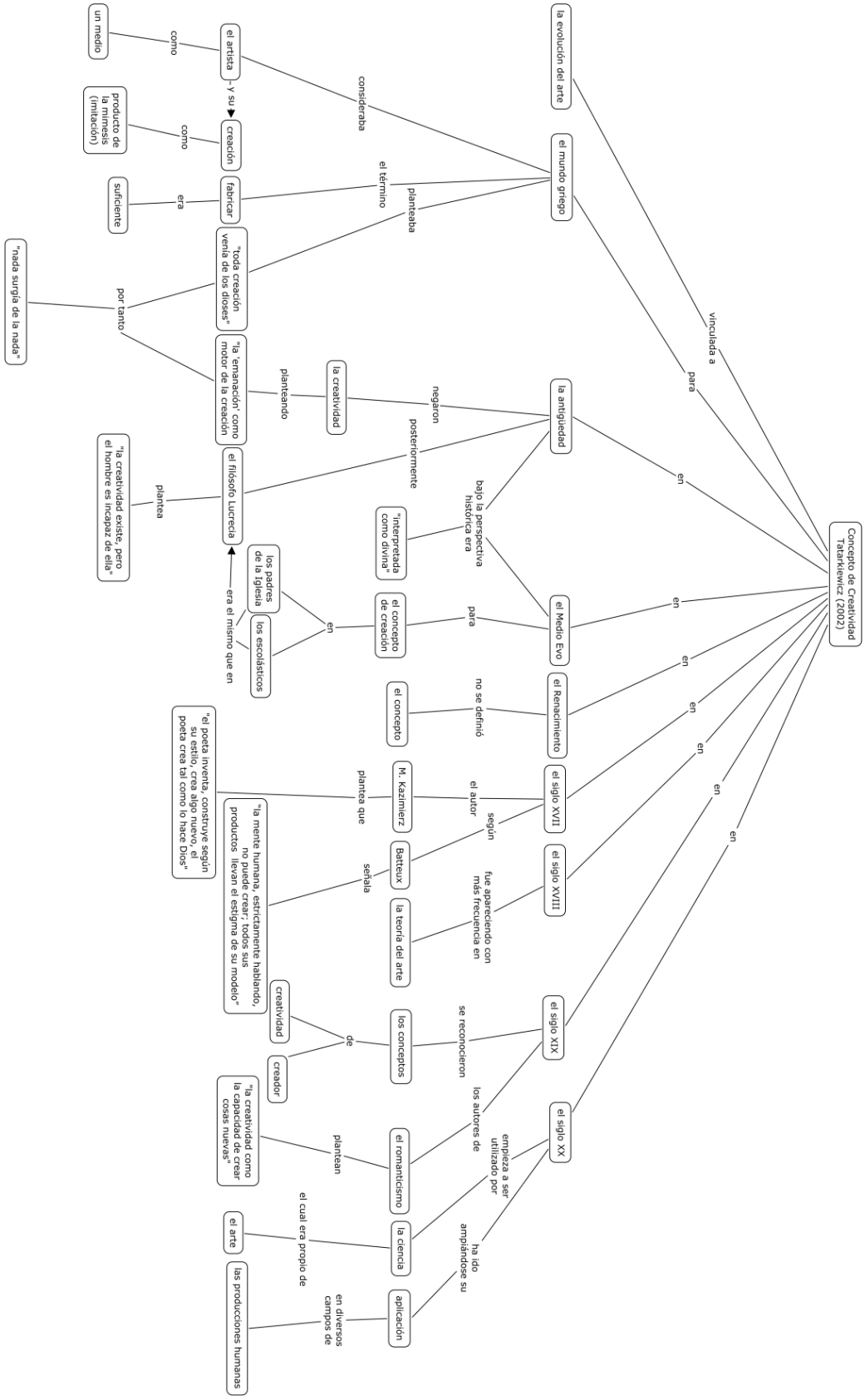

Ilustración 1: Mapa conceptual de la evolución histórica del concepto de creatividad

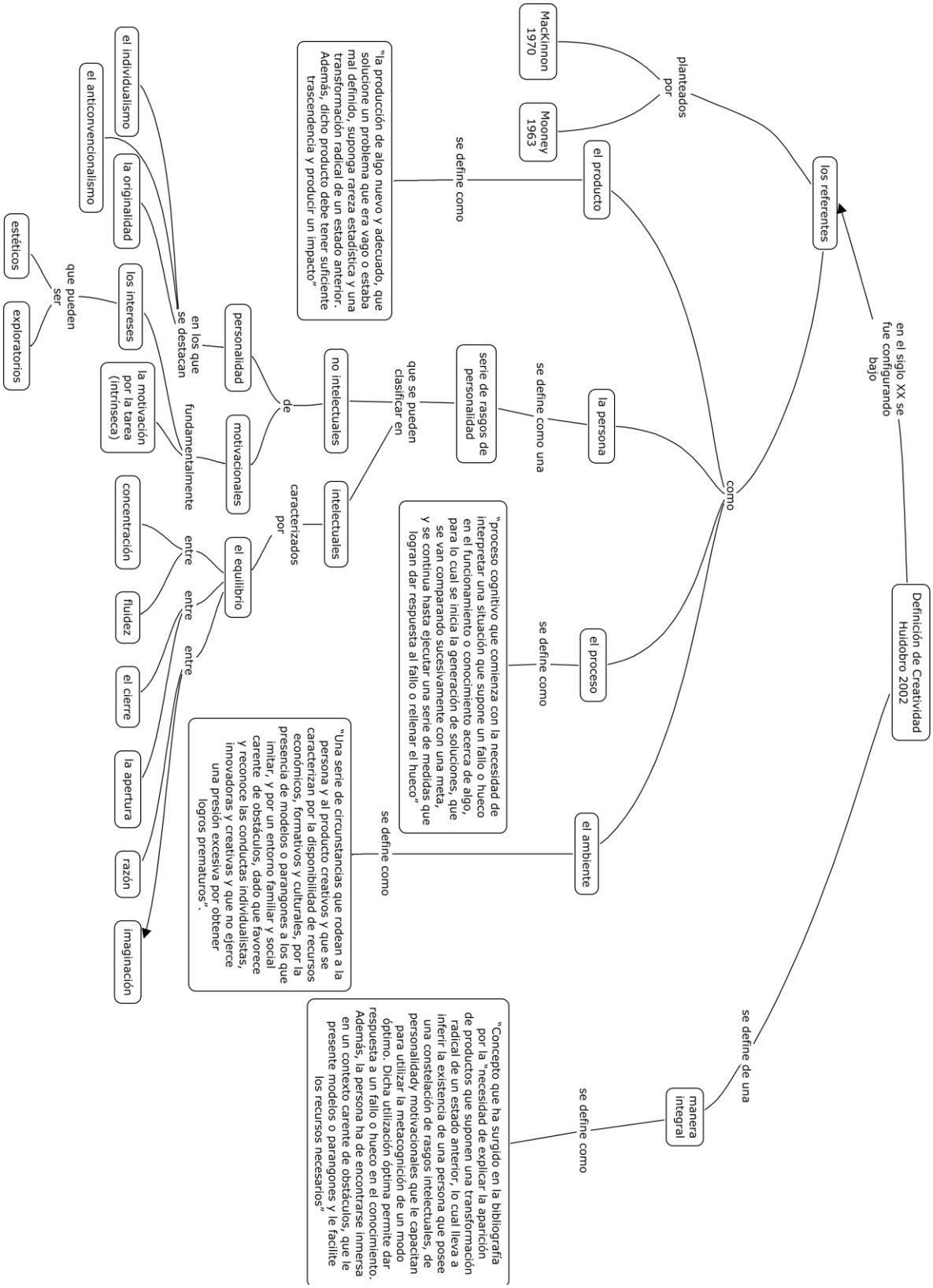

Ilustración 2: Mapa conceptual de la definición de creatividad según Huidobro, 2002