



TRABAJO FIN DE MÁSTER

Análisis y optimización de la representación gráfica de Shape Expressions

*Analysis and optimization of graphical representations of Shape
Expressions*

Jorge Álvarez Fidalgo

DIRECTOR

Jose Emilio Labra Gayo

Máster Universitario en Ingeniería Web - Investigación

Junio 2022

Resumen

Las Shape Expressions (ShEx) se utilizan en diversas ramas del conocimiento para definir estructuras de grafos RDF. Las visualizaciones ShEx permiten a todo tipo de usuarios comprender mejor los esquemas subyacentes y percibir sus propiedades. Sin embargo, el único antecedente (*RDFShape*) adolece de una escalabilidad limitada que perjudica la comprensión en casos de gran tamaño. En este trabajo, se define una notación visual para ShEx construida sobre principios operacionalizados para un diseño cognitivamente eficiente. Además, se implementan dos aproximaciones con mecanismos de gestión de la complejidad a dicha notación: un diagrama 2D (*Shumlex*) y un grafo 3D (*3DShEx*). Se realiza una evaluación comparativa de usuarios entre ambas aproximaciones y *RDFShape*. Los resultados mostraron que los usuarios de *Shumlex* fueron significativamente más rápidos que los de *3DShEx* en esquemas grandes. Aunque no se observaron diferencias significativas en cuanto a las tasas de éxito y la precisión, sólo *Shumlex* obtuvo la máxima puntuación en ambas. Además, aunque las valoraciones de los usuarios fueron mayoritariamente positivas para todas las herramientas, su retroalimentación fue más favorable para *Shumlex*. Por el contrario, la escalabilidad de *RDFShape* y *3DShEx* recibió numerosas críticas. Teniendo en cuenta estos resultados, se concluye que *Shumlex* puede tener potencial como visualización cognitivamente eficiente de *ShEx*. En cambio, la interacción más intrincada que supone el entorno 3D parece ser detrimental para los usuarios de *3DShEx*.

Palabras clave

Shape Expressions, notación visual, UML, 3D, carga cognitiva.

El autor declara la originalidad de los contenidos del presente documento, así como la debida citación de las fuentes utilizadas.

Índice de contenido

1	Introducción.....	8
2	Ejemplo motivador.....	10
3	Trabajo relacionado	11
3.1	Validación de RDF	11
3.2	Aspectos cognitivos del diseño de notaciones visuales	12
3.3	Visualizaciones en la web semántica.....	13
4	Propuesta.....	15
4.1	Notación visual.....	15
4.1.1	Claridad semiótica	16
4.1.2	Discriminabilidad perceptiva.....	16
4.1.3	Transparencia semántica.....	17
4.1.4	Gestión de la complejidad.....	17
4.1.5	Integración cognitiva.....	17
4.1.6	Expresividad visual.....	18
4.1.7	Codificación dual	18
4.1.8	Economía gráfica.....	18
4.1.9	Ajuste cognitivo	18
5	Implementación.....	19
5.1	Shumlex.....	19
5.2	3DShex	21
6	Evaluación	24
6.1	Metodología.....	24
6.1.1	Procedimiento	24
6.1.2	Muestra.....	25
6.1.3	Casos de prueba.....	25
6.1.4	Amenazas a la validez.....	26
6.1.5	Análisis	26

6.2	Resultados	27
6.3	Discusión	29
7	Conclusiones y trabajo futuro	32
8	Dirección y gestión del TFM.....	34
8.1	Planificación del proyecto.....	34
8.1.1	Identificación de interesados	34
8.1.2	OBS, PBS	34
8.1.3	Planificación inicial. WBS.....	35
8.1.4	Riesgos.....	39
8.1.5	Presupuesto inicial.....	42
8.2	Ejecución del proyecto.....	48
8.2.1	Plan de seguimiento de la Planificación.....	48
8.2.2	Bitácora de incidencias del proyecto	48
8.2.3	Riesgos.....	49
8.3	Cierre del proyecto	56
8.3.1	Planificación final	56
8.3.2	Informe final de riesgos.....	59
8.3.3	Presupuesto final de costes	59
8.3.4	Presupuesto final de cliente	64
8.3.5	Informe de lecciones aprendidas.....	65
9	Referencias.....	66
10	Anexos.....	69
10.1	Métrica de similitud	69
10.1.1	Distancia visual.....	69
10.1.2	Codificación redundante	70
10.1.3	Pop-out perceptual	70
10.1.4	Diferenciación textual	71
10.2	Resultados del análisis cualitativo.....	72

10.3	Plan de gestión de riesgos.....	73
10.3.1	Metodología	73
10.3.2	Herramientas y tecnologías	73
10.3.3	Roles y responsabilidades.....	74
10.3.4	Presupuesto.....	74
10.3.5	Calendario.....	74
10.3.6	Categorías de riesgo	75
10.3.7	Definiciones de probabilidad	76
10.3.8	Definiciones de impacto por objetivos.....	76
10.3.9	Matriz de probabilidad e impacto.....	77
10.3.10	Planes de contingencia.....	77
10.3.11	Formatos de la documentación	78
10.3.12	Seguimiento	78
10.4	Artículo de investigación	79

Índice de figuras

Figura 1. Visualización del esquema Genewiki en RDFSShape	10
Figura 2. Arquitectura de Shumlex	20
Figura 3. Visualización del esquema de Genewiki en Shumlex	20
Figura 4. Visualización del esquema de Genewiki con complejidad reducida en Shumlex	21
Figura 5. Arquitectura de 3DShEx.....	22
Figura 6. Visualización del esquema de Genewiki en 3DShEx.....	23
Figura 7. Visualización del esquema de 3DShEx con complejidad reducida en 3DShEx.	23
Figura 8. Precisión en el segundo caso según la herramienta.....	30
Figura 9. PBS del proyecto.	35
Figura 10. Diagrama de Gantt de la planificación inicial	35
Figura 11. Diagrama de Gantt de la planificación final.....	56

Índice de tablas

Tabla 1. Notación visual propuesta.....	15
Tabla 2. Estadísticos descriptivos para los resultados del primer caso de prueba.....	27
Tabla 3. Estadísticos descriptivos para los resultados del segundo caso de prueba. ..	28
Tabla 4. Lista de interesados.....	34
Tabla 5. Recursos del OBS.....	34
Tabla 6. Planificación inicial	36
Tabla 7. Riesgos identificados	39
Tabla 8. Registro de riesgos.....	40
Tabla 9. Partida de Definición	42
Tabla 10. Partida de Estado del arte	42
Tabla 11. Partida de Descripción del sistema.....	43
Tabla 12. Partida de Elaboración de prototipos.....	43
Tabla 13. Partida de Evaluación	44
Tabla 14. Partida de Artículo	44
Tabla 15. Partida de Memoria.....	45
Tabla 16. Partida de Costes indirectos.....	45
Tabla 17. Presupuesto de costes.....	46
Tabla 18. Presupuesto de cliente.....	47
Tabla 19. Bitácora de incidencias.....	48
Tabla 20. Seguimiento de riesgos.....	49
Tabla 21. Hoja del riesgo 1.....	50
Tabla 22. Hoja del riesgo 2.....	52
Tabla 23. Hoja del riesgo 5.....	53
Tabla 24. Hoja del riesgo 9.....	54
Tabla 25. Hoja del riesgo 13.....	55
Tabla 26. Planificación final.....	57
Tabla 27. Partida final de Definición.....	59
Tabla 28. Partida final de Estado del arte	59
Tabla 29. Partida final de Propuesta.....	60
Tabla 30. Partida final de Implementación	60
Tabla 31. Partida final de Evaluación	61
Tabla 32. Partida final de Artículo	61
Tabla 33. Partida final de Memoria	61
Tabla 34. Partida final de Costes indirectos.....	62

Tabla 35. Presupuesto de costes final	63
Tabla 36. Presupuesto de cliente final	64
Tabla 37. Distancia visual para los pares de símbolos gráficos.	70
Tabla 38. Codificación redundante para los pares de símbolos gráficos.	70
Tabla 39. Estadísticos descriptivos de los resultados del análisis cualitativo	72
Tabla 40. Presupuesto asignado a la Gestión de Riesgos	74
Tabla 41. Definiciones de probabilidad.....	76
Tabla 42. Impacto sobre los objetivos principales.....	76
Tabla 43. Matriz de probabilidad e impacto.....	77

1 Introducción

El concepto de “web semántica” fue acuñado en los albores del presente siglo, aludiendo a una nueva generación de la web capaz de ser analizada en su totalidad por ordenadores [1]. Una necesidad imperiosa en una red de tamaño exponencial, cuyo procesamiento supera las capacidades humanas. Con el objeto de satisfacer tal visión, varias tecnologías surgieron como parte de una implementación multicapa de la web semántica, aportando diferentes particularidades: XML, RDF u OWL, entre muchos otros.

En el contexto de dicho ecosistema, en 2014 se propuso *Shape Expressions* (ShEx) [2] como lenguaje para la validación de datos RDF. Mediante la definición de estructuras de grafos RDF, permite a los productores y consumidores de datos acordar una solución común y evitar incoherencias. Dado que RDF reúne a usuarios de diversas ramas del conocimiento humano, ShEx se emplea en una serie de contextos de variada naturaleza. Por ejemplo, se utiliza para validar la representación RDF de FHIR¹, un estándar para el intercambio de datos sanitarios.

Esto implica que sus usuarios no han de estar necesariamente familiarizados con los lenguajes de programación textual, lo que puede suponer una ardua curva de aprendizaje. Una posible solución a este problema reside en el uso de **visualizaciones**. Estas permiten a los usuarios comprender grandes cantidades de datos de manera eficiente y permiten una mejor percepción de las propiedades emergentes, errores y patrones [3].

El único precedente en cuanto a ShEx refiere es *RDFShape* [4], el cual es capaz de generar visualizaciones al estilo de los diagramas de clases de UML² para un subconjunto del lenguaje. Sin embargo, adolece de una escalabilidad limitada, así como de un grado de sobrecarga de símbolos que puede afectar a su transparencia semántica. Por lo tanto, la información transmitida puede ser cognitivamente ineficiente, en particular en los casos de uso más grandes.

Otras visualizaciones de la web semántica formulan diferentes soluciones al problema de desarrollar una notación visual inteligible para los usuarios, con distintos niveles de

¹ <https://www.hl7.org/fhir/rdf.html>

² <https://www.omg.org/spec/UML/>

éxito. Sin embargo, la mencionada cuestión de la escalabilidad -también denominada **gestión de la complejidad**- rara vez se aborda. A lo sumo, se proporcionan mecanismos automáticos para reducir el número de elementos mostrados, pero no se da opción alguna al usuario sobre los detalles.

Así, la principal **contribución** de este trabajo radica en la propuesta de una notación visual para ShEx que pretende ser cognitivamente eficiente -con énfasis en una gestión de la complejidad operada por el usuario-, analizando las implicaciones perceptivas de su materialización tanto en un plano 2D como en un espacio 3D.

El resto del documento está estructurado del siguiente modo. En la Sección 2 se ofrece un ejemplo motivador. En la Sección 3, se proporciona información sobre la validación de RDF, las implicaciones cognitivas del diseño de notaciones visuales y el estado del arte sobre las herramientas de visualización en la Web Semántica. La notación visual propuesta y las dos aproximaciones a la misma se exponen en la Sección 4. La implementación de los prototipos correspondientes se discute en la Sección 5. La metodología de evaluación mediante usuarios, los resultados y su discusión se exponen en la Sección 6. Las conclusiones y el trabajo futuro se discuten en la Sección 7. La dirección y gestión del proyecto se exponen en la Sección 8. La Sección 9 está dedicada a las referencias bibliográficas citadas a lo largo de este trabajo. Por último, la Sección 10 comprende los anexos.

2 Ejemplo motivador

El proyecto *GeneWiki* [5] aspira a utilizar Wikidata como marco semántico para gestionar y difundir datos biomédicos. Para ello, describe un grafo de conocimiento sobre dichas entidades y sus relaciones. Nuestro ejemplo motivador de ShEx³ define dicha estructura. Dado su abundante número de elementos, su correcta visualización supone un desafío técnico.

Su representación visual en *RDFShape* (DOT)⁴ genera un diagrama de clases con 23 clases y más de 70 relaciones (ver Fig. 1). Además de las implicaciones cognitivas que supone procesar un grafo de gran tamaño -que se discutirán más adelante-, podemos apreciar problemas típicos de escalabilidad. Las secciones centrales del diagrama se saturan con relaciones entre *shapes*, cuyos detalles resultan difíciles de discernir. Por lo tanto, es un campo de pruebas adecuado para comprobar los mecanismos de gestión de la complejidad y la eficiencia cognitiva de la notación visual.

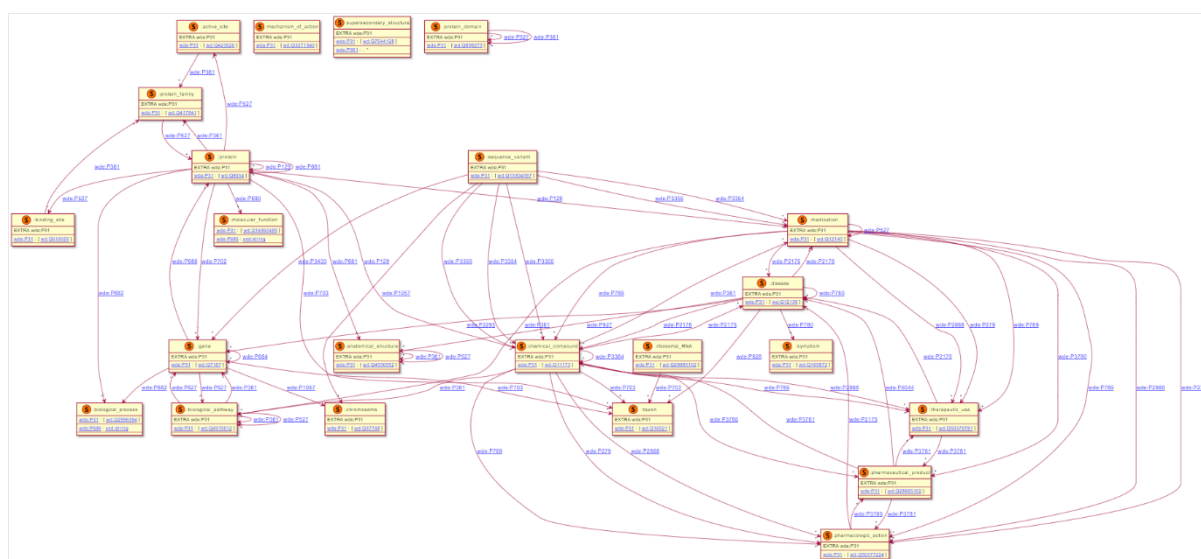


Figura 1. Visualización del esquema Genewiki en RDFShape

³ <https://github.com/weso/sparkwds/sub/blob/master/examples/genewiki.shex>

⁴ <https://rdfshape.weso.es/link/16520062624>

3 Trabajo relacionado

En esta sección, exponemos (i) una introducción a la validación de RDF, (ii) las implicaciones cognitivas del diseño de notaciones visuales y (iii) visualizaciones en la web semántica.

3.1 Validación de RDF

RDF es un modelo estándar para el intercambio de datos en la Web⁵. Se basa en el concepto de tripletas. Cada triplete comprende un sujeto, un predicado y un objeto. Por tanto, define una relación (propiedad) entre el recurso indicado por el primero y el segundo, que puede ser un recurso o un valor literal. Un recurso puede ser cualquier ente referido por un IRI. Una triplete puede ilustrarse como un arco dirigido entre dos vértices; por lo tanto, se puede componer fácilmente un grafo uniendo todas las tripletas, lo cual se trata de una expresión típica de los datos RDF.

Shape Expressions (ShEx) es un lenguaje de esquemas que describe tales estructuras de grafos RDF [2]. Fue creado con el objetivo de proporcionar una herramienta de validación y transformación para RDF, en un rol similar al de los lenguajes de esquema en XML. Así, podemos describir un esquema ShEx y comprobar la conformidad de los nodos de un grafo RDF con dicho esquema.

Shapes Constraint Language (SHACL)⁶ es un lenguaje para la validación de grafos RDF mediante *shapes* y otras construcciones, publicado como recomendación del W3C. Por lo tanto, comparte un objetivo común con ShEx, así como un conjunto de características, por ejemplo: *shapes*, restricciones de nodos, restricciones de propiedades y operadores lógicos. Sin embargo, existen discrepancias sustanciales en sus respectivos diseños que los diferencian. ShEx pretende describir estructuras de grafos RDF, mientras que SHACL pretende proporcionar un conjunto de restricciones a satisfacer. En consecuencia, SHACL emite información detallada cuando las restricciones no se cumplen, pero ofrece escasa información cuando no se producen errores. Por el contrario, las implementaciones de ShEx informan sobre el gráfico validado, incluyendo pruebas de la conformidad de sus nodos, mientras que son algo opacas respecto a los casos no válidos [6].

⁵ <http://www.w3.org/RDF/>

⁶ <https://www.w3.org/TR/shacl/>

3.2 Aspectos cognitivos del diseño de notaciones visuales

La **teoría de la carga cognitiva** “se ocupa de la forma en que los recursos cognitivos se centran y utilizan durante el aprendizaje y la resolución de problemas” [7]. Describe el deterioro de la comprensión que se produce cuando los procesos de aprendizaje conducen a esfuerzos cognitivos ulteriores. Se realiza una distinción entre carga cognitiva intrínseca y extrínseca; la primera se debe a la complejidad inherente de la información, mientras que la segunda se genera debido a la forma en que se presenta dicha información. Estos fenómenos están relacionados de tal manera que las consecuencias de la carga cognitiva extrínseca sólo pueden ser perceptibles cuando tiene lugar una carga cognitiva intrínseca causada por la alta interactividad de los elementos [8].

D. Moody describe en su teoría de la **física de las notaciones** (PoN) una serie de principios para diseñar notaciones visuales cognitivamente eficaces [9]: *claridad semiótica*, *discriminabilidad perceptiva*, *transparencia semántica*, *gestión de la complejidad* (reducción de la carga cognitiva extrínseca), *integración cognitiva*, *expresividad visual*, *codificación dual*, *economía gráfica y ajuste cognitivo*. En la última década, la PoN se ha convertido en un estándar ampliamente utilizado para el diseño de notaciones en detrimento de sus competidores [10].

Se han formulado varias críticas a la PoN. La operacionalización de dichos principios va desde medidas objetivas -la claridad semiótica es una correspondencia 1:1- hasta evaluaciones subjetivas -la “sugerencia de significado” que implica la transparencia semántica sólo puede determinarse por medios empíricos- [11]. Esto implica un grado de implicación del usuario del que suele carecer su aplicación [12]. Se han planteado propuestas posteriores con el objeto de mejorar dicha operacionalización, ya sea parcial [13] o completamente [14].

En otro orden de cosas, el impacto de las visualizaciones 3D en la carga cognitiva podría estar estrechamente relacionado con la capacidad espacial [15]. Es decir, los sujetos con alta capacidad espacial perciben su carga cognitiva como baja y viceversa. Otras investigaciones muestran que este efecto se agrava cuando se trata de visualizaciones estáticas; las interacciones dinámicas generan un efecto compensatorio para los usuarios con baja capacidad espacial [16].

3.3 Visualizaciones en la web semántica

Como se indicó en la introducción, RDFSShape proporciona la única visualización disponible actualmente para *Shape Expressions*⁷. Esta genera un gráfico bidimensional en el que las cajas de UML simbolizan *shapes* y las flechas direccionales representan referencias a otras *shapes*. No se proporcionan acciones interactivas ni mecanismos de gestión de la complejidad.

Se han llevado a cabo proyectos de similar índole para el estrechamente relacionado SHACL en forma de editores visuales. Arndt *et al.* implementaron una herramienta basada en Ontopad⁸ que permite componer un modelo de datos visual de SHACL [17]. La mayor parte de la interacción se realiza a través de una interfaz textual; los nuevos elementos deben ser arrastrados a un lienzo para formar parte de la visualización. Los usuarios pueden realizar unas pocas tareas sobre la visualización, como la vinculación de propiedades.

Lieber *et al.* definen sendas notaciones visuales basadas en UML y VOWL para representar restricciones RDF y las implementan en *UnSHACLed*, un editor visual para SHACL [18]. Las pruebas empíricas no mostraron diferencias significativas en las tasas de error entre las dos aproximaciones. No obstante, la mayoría de los usuarios prefirió la notación basada en VOWL. Los autores reconocen la necesidad de mecanismos de gestión de la complejidad, pero lo consideran fuera de su alcance.

VOWL [19] es una notación visual para representar ontologías OWL⁹, con dos implementaciones disponibles: *WebVOWL* y *ProtégéVOWL*. *WebVOWL* hace uso de un grafo de fuerza, lo que permite la interacción del usuario con el posicionamiento de los elementos. Además, ofrece una herramienta de gestión de la complejidad: una función de colapso que reduce el número de elementos en pantalla, si bien no otorga al usuario la posibilidad de elegir los detalles. No obstante, esto no impide que se produzca una superposición de elementos cuando existe un gran número de relaciones entre dos nodos.

Dicha superposición ha sido un problema recurrente en la historia de las visualizaciones en la web semántica. Ya en el año 2000 surgió una posible solución: representar los

⁷ <https://rdfshape.weso.es/shexInfo>

⁸ <https://github.com/AKSW/OntoPad>

⁹ <https://www.w3.org/TR/owl-features/>

grafos semánticos en tres dimensiones. Con tal propósito, herramientas como *UNIVIT* [20] y *NV3D* fueron implementadas; desafortunadamente, sus notaciones visuales eran demasiado dependientes de la combinación arbitraria de variables visuales (forma, color...) para transmitir eficientemente datos complejos [21].

Una década más tarde, surge *X3D-UML* como una implementación de UML en 3D, particularmente centrada en diagramas de máquinas de estado [22]. Consiste en una serie de planos interconectados en un espacio tridimensional, cada uno de los cuales muestra un diagrama UML 2D. Por lo tanto, se trata más bien de una solución intermedia.

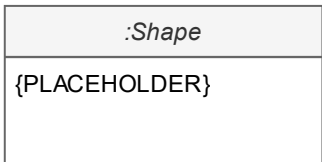
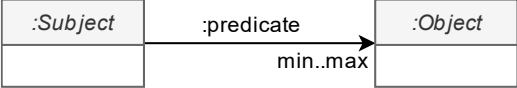
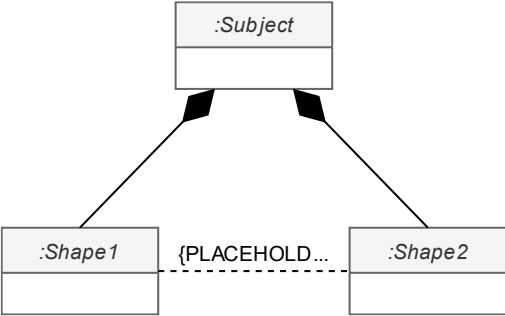
4 Propuesta

Se propone una notación visual inspirada en un diagrama de clases UML para representar gráficamente ShEx, con dos aproximaciones alternativas: un diagrama 2D (*Shumlex*) y un grafo dirigido 3D (*3DShEx*).

4.1 Notación visual

La notación visual propuesta se muestra en la Tabla 1. Su diseño está estructurado según los principios de la PoN, expuestos a continuación.

Tabla 1. Notación visual propuesta.

Nombre	{PLACEHOLDER}	Representación gráfica	Ejemplo
Triple Constraint	<Property> <NodeConstraint> <Cardinality>;		:User { :name xsd:string ?; }
EachOf	<TripleConstraint>+		:User { :name xsd:string; :age xsd:integer; }
Top-level Nodekind	nodeKind: <NodeKind>		:HomePage IRI
Extra	EXTRA <ValueSet>		:User EXTRA :name { :name xsd:string;}
Closure	CLOSED		:User CLOSED { }
ShapeRef			
ShapeAnd	AND		:User { :name xsd:string; } AND { :age xsd:integer; }
ShapeOr	OR		:User { :name xsd:string; } OR { :age xsd:integer; }
OneOf	OneOf		:User {

			:name xsd:string; :givenName xsd:string; }
ShapeNot	NOT		:NoName NOT { schema:name: . }
Labelled	Composed Of	<pre> classDiagram class Subject class Shape Subject "1" *-- "1" Shape : {PLACEHOLDER} </pre>	:User { \$.name (:name .; :email IRI;); }

4.1.1 Claridad semiótica

La claridad semiótica se sacrifica en favor de la transparencia semántica y la economía gráfica. En primer lugar, la notación incurre en un caso deliberado de **déficit simbólico**, ya que existen una serie de constructos semánticos sin una representación gráfica asignada a ella. Las restricciones de los nodos se muestran textualmente dentro de las *shapes*, al modo de los atributos en las clases UML. De este modo, esperamos aprovechar la popularidad de los diagramas de clase UML para transmitir la información a un público más amplio (ver 4.1.3).

En segundo lugar, algunas construcciones semánticas emplean la misma construcción visual (**sobrecarga simbólica**) con diferenciación textual. Dado que esas construcciones semánticas son conceptualmente similares (por ejemplo, la conjunción y la disyunción), esperamos lograr una economía gráfica sin perturbar su claridad.

4.1.2 Discriminabilidad perceptiva

Para determinar objetivamente la facilidad de discriminación entre símbolos, es necesario definir tanto una métrica como un umbral de similitud y un umbral de disimilitud entre dos símbolos gráficos [14]. Para ello, modificamos ligeramente la propuesta de [13] para calcular la media de los siguientes criterios: distancia visual (VD), codificación redundante (RC), *pop-out* perceptual (PPO) y diferenciación textual (TD). Estos se normalizan en un intervalo de [0, 1], de forma que 0 denote una discriminabilidad nula y 1 el cumplimiento de todos los criterios. Se elige 0,5 como umbral de disimilitud.

En aras de la brevedad, los detalles de las modificaciones y los cálculos se exponen en el Anexo 10.1. Los valores obtenidos fueron VD = 0,47, RC = 0,29, PPO = 1 y TD =

0,5. Esto da como resultado un valor medio de 0,57, lo que demuestra la discriminabilidad perceptiva positiva.

4.1.3 *Transparencia semántica*

Al dar forma a la notación de tal modo que se asemeje a un diagrama de clases UML, el objetivo es aumentar su transparencia semántica, especialmente para los usuarios novicios. Cuanto menos, se espera conseguir translucidez semántica, de modo que las construcciones visuales proporcionen una pista sobre su significado por asociación. No obstante, esto no se puede verificar hasta que se realice una evaluación por parte del usuario, dada su mencionada subjetividad.

4.1.4 *Gestión de la complejidad*

La gestión de la complejidad se aborda de diferentes maneras. *Shumlex* se inspira en la utilidad de modularización implementada en *GraphQL Voyager*¹⁰, que al hacer clic en una de las clases resalta únicamente esa clase y sus relaciones con sus vecinas, reduciendo drásticamente la visibilidad de las demás. Proponemos una modificación de este concepto para que sea acumulativo, es decir, que al hacer clic en una segunda clase no se cambie el foco a esa clase, sino que se añada al conjunto resaltado.

En lo que respecta a 3DShEx, se implementará un *zoom* inteligente [23]. Esto implica una interactividad adicional además del *zoom* común. Por defecto, los nodos sólo mostrarán el identificador de la *shape*. Cuando se interactúe con ella (“abrir”), los nodos se expandirán para revelar las restricciones pertinentes (*caja negra*). La interacción posterior (“cerrar”) lo devolverá a su estado inicial. Como herramienta adicional de gestión de la complejidad, se propone una sencilla función de colapso que, bajo petición, muestra sólo el nodo deseado y sus vecinos. Ambas proporcionan una capa de abstracción.

4.1.5 *Integración cognitiva*

Dado que no se utilizan múltiples diagramas para representar un único conjunto de datos, este principio no se aplica.

¹⁰ <https://apis.guru/graphql-voyager/>

4.1.6 *Expresividad visual*

La PoN se basa en la lista de variables visuales de Bertin [24]: forma, textura, brillo, tamaño, color, orientación y variables planas. La notación propuesta utiliza las siguientes: forma, textura y brillo. Por lo tanto, se encuentra en un punto intermedio entre la unidimensionalidad visual y la saturación visual. Moody afirma que la mayoría de los diagramas en la ingeniería de software son visualmente unidimensionales [9] y, por lo tanto, logramos un mayor grado de discriminabilidad perceptiva que aquellos, en función de lo expuesto anteriormente.

4.1.7 *Codificación dual*

Cada representación visual se complementa con un texto que proporciona una pista sobre su significado.

4.1.8 *Economía gráfica*

Como ya se ha detallado, el déficit simbólico fue introducido *ex profeso* para reducir la complejidad gráfica. En consecuencia, el número de símbolos gráficos es de 4. Se trata de un valor inferior al límite de capacidad de procesamiento de 7 ± 2 propuesto por Miller [25], sugerido por van der Linden *et al.* [14] como referencia para este principio.

4.1.9 *Ajuste cognitivo*

El objetivo de la mencionada aproximación dual es maximizar el ajuste cognitivo, priorizando cada uno diferentes necesidades. *Shumlex* pretende ser más próximo a la especificación del diagrama de clases UML, con la intención de ser accesible a un público más amplio no necesariamente familiarizado con los detalles técnicos. Así, toda la información está inicialmente disponible como lo estaría en un diagrama común, lo que hace que la visualización esté más centrada en las restricciones en contraste con un *3DShex* con énfasis en las relaciones.

Por el contrario, *3DShex* tiene un carácter más experimental, que al presentar de forma interactiva la misma información en un espacio tridimensional pretende analizar más a fondo las implicaciones cognitivas en su audiencia y los posibles beneficios que pueda aportar. Los detalles de las *shapes* se ocultan tras una capa de abstracción, dando así mayor importancia al diagrama en su conjunto.

5 Implementación

En esta sección, describimos la elaboración de los prototipos para ambas aproximaciones a la notación visual.

5.1 Shumlex

Para construir la visualización hacemos uso de *Mermaid*¹¹, una librería de *Javascript* para la generación de diagramas basados en texto. La arquitectura del prototipo se ilustra en la Fig. 2 y consiste en lo siguiente:

1. Un motor de **conversión** que recibe una entrada ShEx y genera la sintaxis tipo *Markdown* que requiere *Mermaid*. Dado que esta librería no acepta una serie de símbolos utilizados en ShEx, es necesario utilizar alternativas. Por ejemplo, no se permite el uso de dos puntos; el término prefijado “:Usuario” tendría que codificarse como “_Usuario”.
2. Un módulo de **visualización** que invoca la librería con el resultado anterior para generar un SVG. Una vez visualizado, los textos saneados son sustituidos por los originales.
3. Un módulo de **post-procesamiento** que implementa la funcionalidad de gestión de la complejidad. Asigna a cada clase del diagrama un evento que, al hacer clic, reduce al mínimo la opacidad de todos los elementos excepto esa misma clase, sus relaciones y los objetivos de éstas. Existen un par de excepciones: a) no oscurecerá los elementos ya resaltados y b) revertirá el efecto si ya se ha aplicado a dicha clase. Además, al pasar el ratón por cualquier etiqueta se comprobará la existencia de la entidad en Wikidata y se mostrará su significado mediante un *tooltip*. El propósito de esto es aumentar la comprensión de las expresiones de forma comúnmente usadas y relacionadas con Wikidata, en las que las formas y los predicados son semánticamente opacos (por ejemplo, *wd:Q42944* refiere al CERN).

¹¹ <https://mermaid-js.github.io/mermaid/>

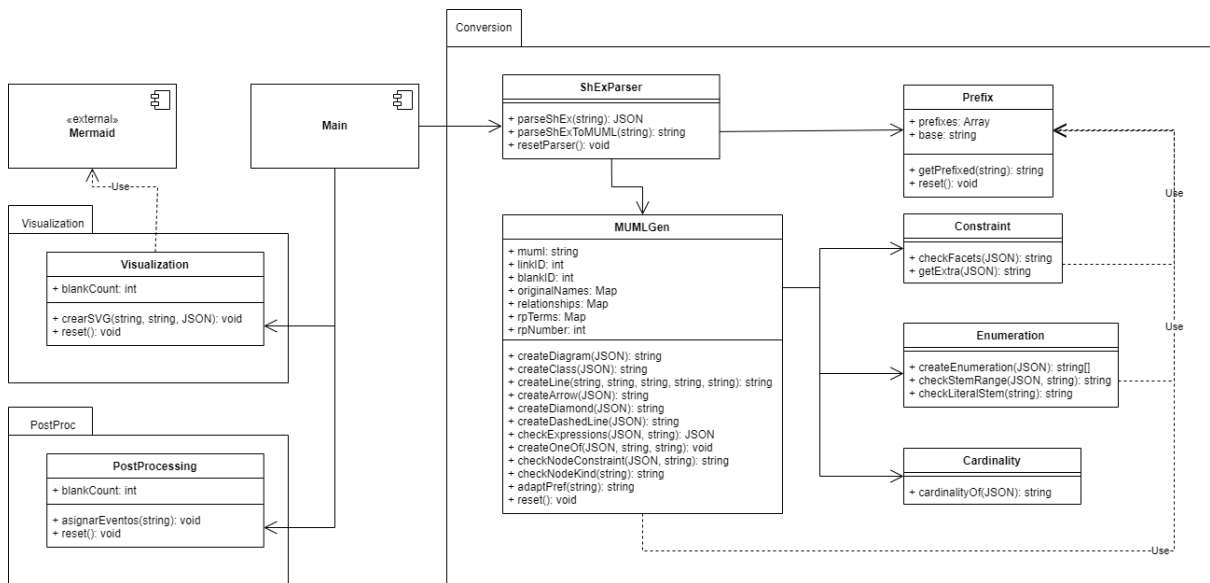


Figura 2. Arquitectura de Shumlex

La aplicación de *Shumlex* al ejemplo motivador se muestra en la Fig. 3. A pesar de que las relaciones están más espaciadas que en la visualización de *RDFShape*, las áreas con alta concentración de elementos siguen estando cognitivamente sobrecargadas. Como se muestra en la Fig. 4, el mecanismo de gestión de la complejidad permite una visualización parcial de los componentes deseados. En el ejemplo proporcionado, la atención se centra en *:medicamento*, destacando así sus relaciones con otras *shapes*.

Este prototipo está públicamente disponible en <http://www.weso.es/shumlex/>.

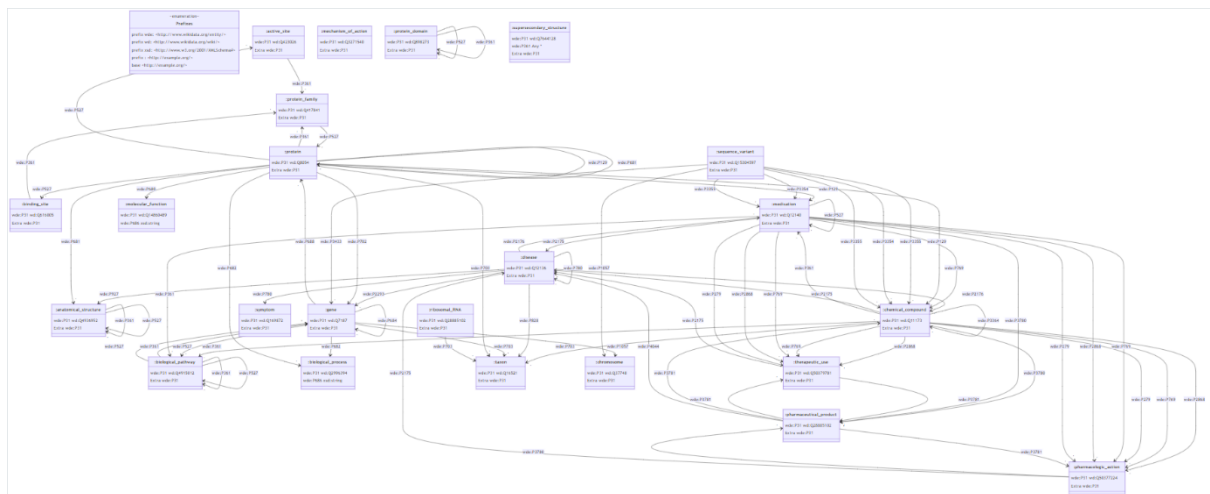


Figura 3. Visualización del esquema de Genewiki en Shumlex

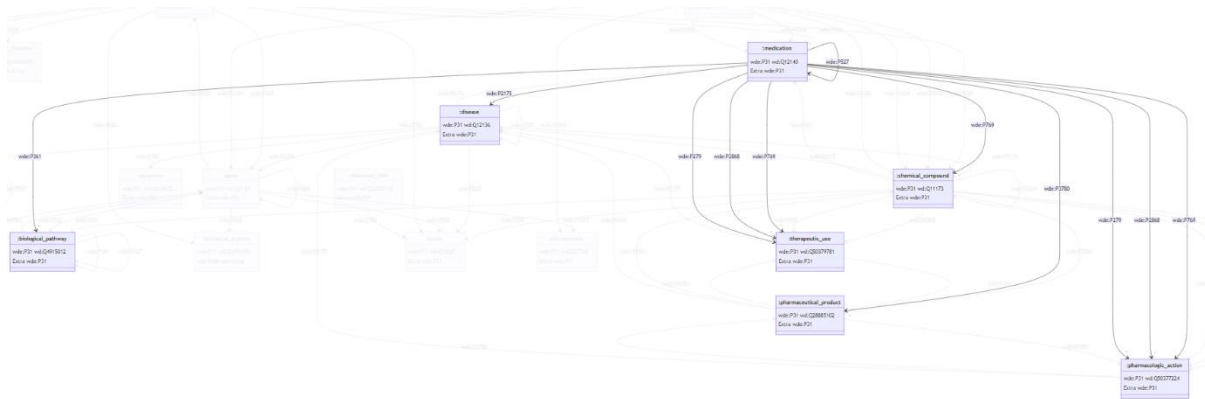


Figura 4. Visualización del esquema de Genewiki con complejidad reducida en Shumlex

5.2 3DShex

Para este prototipo hacemos uso de *3D Force Graph*¹² (3DFG), una librería de *NodeJS* para representar datos gráficos en un espacio tridimensional. La arquitectura del prototipo se expone en la Fig. 5 y consiste en lo siguiente:

1. Un motor de **conversión** que recibe una entrada ShEx y genera los datos JSON que 3DFG requiere. Además de los parámetros necesarios, se incluye información adicional para facilitar la siguiente fase:
 - a. *Lista de restricciones de un nodo.* Información que se mostrará bajo demanda, equivalente a los atributos de clase en UML.
 - b. *Nombre y cardinalidad de una relación.*
 - c. *Curvatura de un enlace.* Los enlaces curvos permiten distinguir las relaciones entre un par de nodos, mientras que los enlaces rectos se solapan entre sí. Por lo tanto, las referencias entre *shapes* (*:User :works @:Company*) requieren las primeras, ya que puede haber cualquier número de ellas. Por el contrario, las relaciones de composición (*ShapeAnd*, *ShapeOr*, *OneOf*, *ShapeNot* y *Labelled* en la Tabla 1) son exclusivas de la *shape* de origen y, por tanto, pueden representarse mediante enlaces rectos.
 - d. *Cabeza de flecha.* Como se muestra en la notación, hay tres posibilidades: ninguna, flecha o diamante.
 - e. *Rotación.* Según lo anteriormente mencionado, los enlaces curvados permiten una disposición libre de la superposición descrita en la Sección 3.3. Sin embargo, por defecto, los enlaces se muestran en la misma

¹² <https://github.com/vasturiano/3d-force-graph>

posición. 3DFG permite especificar un valor de rotación -tomando el nodo como centro de una hipotética circunferencia-, pero los cálculos se dejan en manos del usuario. Por lo tanto, para cada par de nodos, registramos todos sus enlaces y dividimos la circunferencia en partes iguales. Además, debemos tener en cuenta el origen del enlace, ya que desde la perspectiva de cada circunferencia el ángulo será diferente para una determinada posición (por ejemplo, π en el nodo origen equivale a 2π en el destino).

2. Un módulo de **visualización** que hace uso de la información anterior para invocar a la librería. Se utilizan objetos HTML para construir los nodos, por lo que podemos personalizar su contenido así como asignarles un comportamiento dinámico (las restricciones están ocultas por defecto). Se habilitan las siguientes funcionalidades:

- Resaltar al pasar el ratón.* Tanto los enlaces como los nodos poseen esta propiedad; en el caso de estos últimos, también se resaltan sus vecinos y las relaciones correspondientes. En los enlaces, se muestran las partículas en movimiento para reforzar su sentido.
- Detalles.* Al hacer clic en un nodo, todas sus restricciones se muestran en un cuadro expandido. Otro clic lo devuelve a su estado original.
- Colapsar.* Al hacer clic con el botón derecho del ratón sobre un nodo se muestra un grafo reducido, compuesto por dicho nodo y sus vecinos.
- Tooltips de Wikidata.* Al igual que en *Shumlex*.

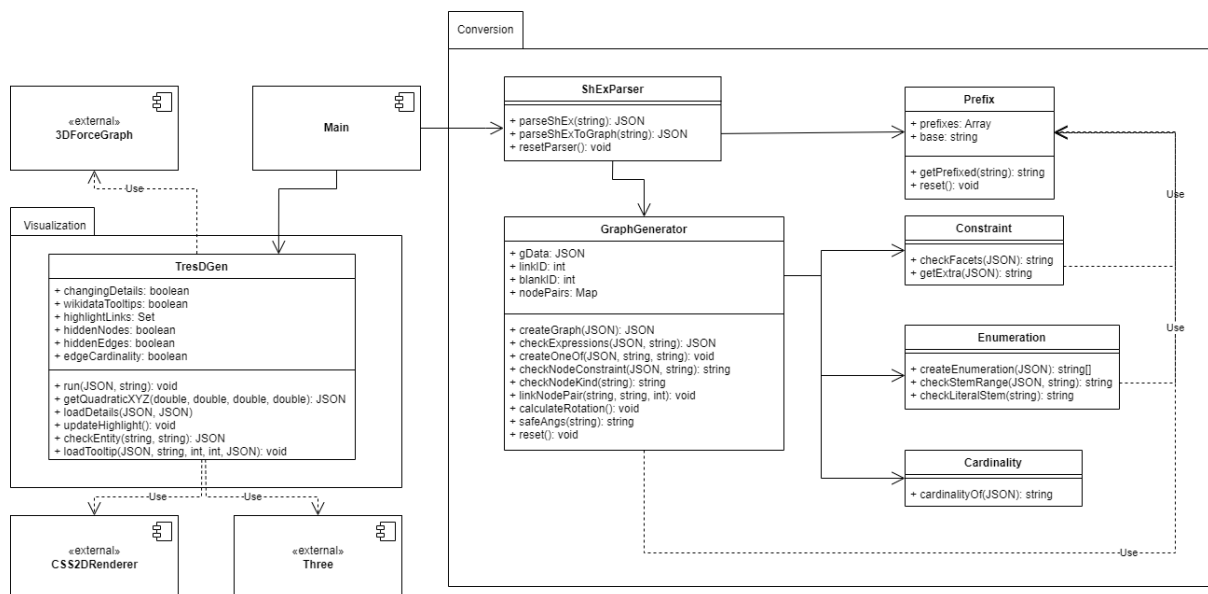


Figura 5. Arquitectura de 3DShEx

La aplicación de *3DShEx* al ejemplo motivador se muestra en la Fig. 6. Aunque las imágenes estáticas resultan de escasa ayuda para su comprensión -ya que puede examinarse desde cualquier posición-, resulta evidente que los clústeres de *shapes* altamente interdependientes exceden los límites de la memoria de trabajo. Así pues, el mecanismo de gestión de la complejidad se aplica de nuevo a *:medicamento*. Como se expone en la Fig. 7, se muestra un grafo mucho más pequeño, compuesto por el elemento deseado y sus vecinos.

El prototipo se encuentra públicamente disponible en <http://www.weso.es/3dshex/>.

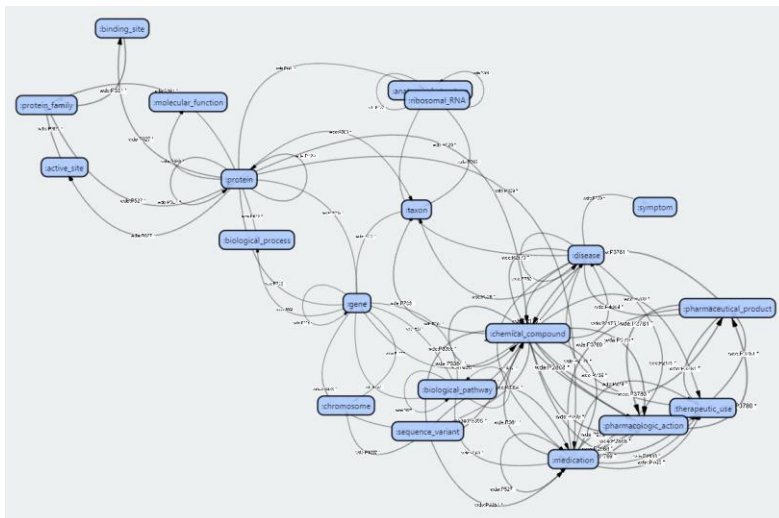


Figura 6. Visualización del esquema de Genewiki en 3DShEx.

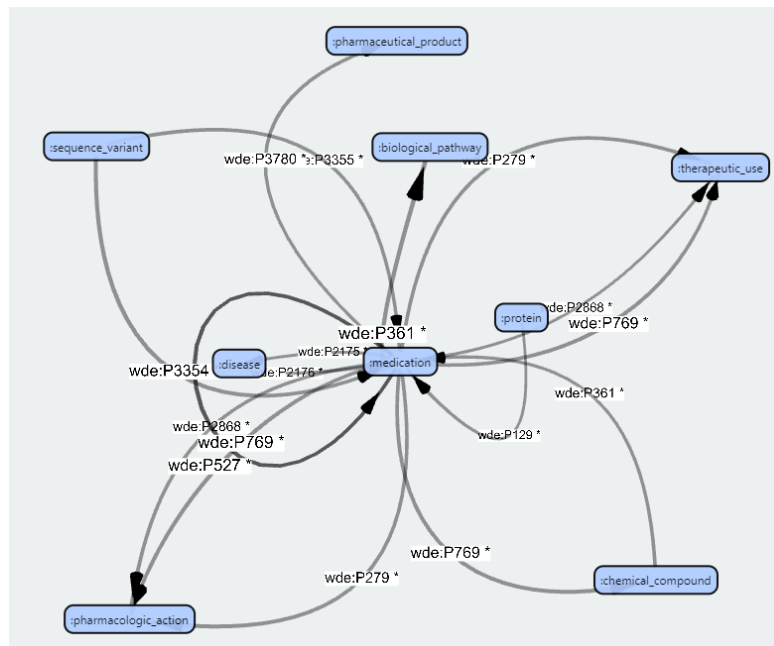


Figura 7. Visualización del esquema de 3DShEx con complejidad reducida en 3DShEx.

6 Evaluación

Con el objeto de probar la notación propuesta, se ha realizado un experimento basado en el realizado en [26]. La metodología empleada, los resultados obtenidos y su discusión se detallan en las siguientes subsecciones. Los conjuntos de datos, los cuestionarios, los manuales y los resultados anonimizados están disponibles libremente en <https://github.com/fidalgoLXXVI/shex-visualization-paper>.

6.1 Metodología

Este estudio de usabilidad sigue un diseño “entre sujetos”, en el que se expone a cada participante a una única herramienta y se le pide que realice una serie de tareas medidas. Sobre los resultados obtenidos se realiza un análisis tanto cuantitativo como cualitativo. A continuación, se analiza con más detalle la metodología de este experimento.

6.1.1 Procedimiento

El experimento se divide en los siguientes pasos:

1. **Cuestionario preliminar.** Se pregunta a los sujetos sobre sus antecedentes y se solicita una autoevaluación de las habilidades relevantes -como experiencia con UML o capacidad espacial-.
2. **Descripción de las herramientas.** Se proporciona a los participantes un breve manual que describe brevemente el funcionamiento y las características de la herramienta correspondiente. Las herramientas seleccionadas para el experimento fueron *RDFShape*, *Shumlex* y *3DShEx*. De este modo, las distintas implementaciones de la notación pueden compararse entre sí, así como con la solución existente.
3. **Cuestionario principal.** Se solicita a los participantes una serie de tareas sobre los casos de prueba. Mediante la herramienta asignada, los sujetos deben intentar realizarlas mientras sus interacciones son medidas por un cronómetro. Cada caso de prueba comprende las siguientes tareas, cuyo objetivo es determinar la capacidad del usuario para navegar por el diagrama y comprender las distintas equivalencias semánticas.
 - a. Buscar una *shape* por su nombre.

- b. Buscar una *shape* con una restricción específica.
 - c. Listar las restricciones intra-nodo de una *shape*.
 - d. Encontrar una referencia entre dos *shapes*.
 - e. Determinar sujeto y objeto de una referencia.
 - f. Listar los vecinos de una *shape*.
4. **Cuestionario final.** Para realizar el análisis cualitativo se formulan a los participantes una serie de preguntas basadas en la escala de Likert. Esto nos permite obtener una serie de variables: *satisfacción general, facilidad de uso, aprendizaje, transparencia semántica, utilidad, propensión al error, escalabilidad, gestión de la complejidad, comprensión de restricciones y comprensión de referencias.*

6.1.2 Muestra

La muestra estaba formada por 13 estudiantes del Máster en Ingeniería Web de la Universidad de Oviedo. Este experimento tuvo lugar en el último día de una asignatura de web semántica, donde se les enseñaron los fundamentos de tecnologías como RDF o ShEx. La mayoría de los participantes comparten una formación demográfica y académica similar, con un grado en Ingeniería Informática del Software.

Según los resultados de la autoevaluación a) el 92,3% tiene un conocimiento medio o alto de UML, b) el 84,6% tiene conocimientos básicos de RDF, c) el 69,2% tiene conocimientos básicos de ShEx mientras que el 15,4% declara no tener conocimientos sobre el tema y d) el 69% tiene una capacidad espacial alta mientras que el resto declara una capacidad espacial media.

6.1.3 Casos de prueba

En este experimento se utilizan dos casos de prueba. El primero se basa en el esquema ShEx de *WebIndex* propuesto en [6], "una de las primeras aplicaciones prácticas de ShEx". Se han realizado algunas modificaciones para reflejar todas las características de las que dispone la notación visual, incluyendo así las operaciones lógicas y la composición. La restricción *OneOf* se ha eliminado, ya que la versión actual de *RDFShape* no la ha implementado todavía. Este esquema presenta unas pocas referencias, centrándose más en otros tipos de restricciones nodales.

El segundo es el esquema de *GeneWiki*, ya descrito en la Sección 2. Tiene aproximadamente el triple de *shapes* que el anterior y más de 70 referencias entre ellas, con escaso número de restricciones de otro tipo.

Por lo tanto, de acuerdo con las implicaciones expuestas en la sección 3.2, los participantes se enfrentarían a distintas cargas cognitivas. El primer caso de prueba tiene una mayor carga intrínseca, dada la mayor complejidad inherente al uso de rasgos semánticos complejos, como las conjunciones y la composición; si bien consta de pocos elementos. Por el contrario, el segundo caso de prueba tiene menor complejidad implícita -la mayoría son simples referencias a otras *shapes*- pero su gran cantidad de elementos provoca una complejidad diagramática -extrínseca- en la visualización.

6.1.4 Amenazas a la validez

Tomando como referencia la lista de amenazas internas y externas a la validez propuesta en [27], se han identificado las siguientes:

- *Selección*. Los participantes pueden compartir ciertas características que les predispongan a obtener los mismos resultados, sobre todo teniendo en cuenta los antecedentes comunes. Para mitigar esta amenaza, los sujetos se distribuyen aleatoriamente entre los grupos experimentales, de modo que esas características puedan estar igualmente distribuidas.
- *Pruebas*. Los participantes pueden familiarizarse con los casos de prueba y recordar las respuestas para tareas posteriores. Para mitigar esto, se tiene especial cuidado en utilizar diferentes fragmentos del esquema y evitar las repeticiones.
- *Interacción de la selección y el tratamiento*. Debido a la limitada variedad de los participantes, la generalización a individuos de otros contextos puede no ser posible. Por tanto, se deben restringir las afirmaciones sobre la universalidad de los resultados. Sin embargo, dada la naturaleza altamente especializada de la contribución, esta cuestión se atenúa.

6.1.5 Análisis

Se recogen y anonimizan sendos resultados cuantitativos y cualitativos. A partir de ellos, se calculan las siguientes variables para cada caso de prueba: **tiempo transcurrido**, **tasa de éxito** y **precisión**. El tiempo transcurrido (T_c) se calcula como la suma del tiempo transcurrido para responder a cada pregunta. La tasa de

éxito (S_c) se calcula como el número de respuestas correctas dividido por el número de preguntas. La precisión (P_c) se calcula como la división del tiempo mínimo transcurrido de todos los participantes por el tiempo transcurrido del estudiante actual, multiplicado por la tasa de éxito. Esta medida da una idea de la rapidez de los participantes a la par que tiene en cuenta su eficacia. Por lo tanto, dado un caso de prueba c y un estudiante sn :

$$P_{csn} := \frac{\min(\{T_{cs1}, \dots, T_{csn}\})}{T_{csn}} \cdot S_{csn}$$

Para el análisis estadístico se emplea *R* 4.2.0. Las comparaciones entre los tres grupos se realizan mediante un ANOVA de una vía siempre y cuando se cumplan los supuestos, eliminando los valores atípicos si fuera necesario. En caso contrario, se utiliza el test de Kruskal-Wallis.

6.2 Resultados

Los estadísticos descriptivos de los resultados cuantitativos para el primer caso de prueba se muestran en la Tabla 2. Las puntuaciones medias de *Shumlex* son sistemáticamente mejores que las de *RDFShape*, y éstas mejores que las de *3DShEx*. Sin embargo, esas diferencias entre los tres grupos no son estadísticamente significativas para ninguna de las variables: $F(2,8)=1.1$; $p=0,377$, $F(2,10)=1,67$; $p=0,236$ y $F(2,9)=1,29$; $p = 0,32$ para T, S y P respectivamente.

Tabla 2. Estadísticos descriptivos para los resultados del primer caso de prueba.

Variable	Grupo	Media	s	max	min
Segundos transcurridos	3DShEx	256,2	66,55	355	210
	RDFShape	210,2	119,62	411	95
	Shumlex	196	95,63	302	73
Tasa de éxito	3DShEx	0,667	0,136	0,833	0,5
	RDFShape	0,7	0,139	0,833	0,5
	Shumlex	0,833	0,136	1	0,667
Precisión	3DShEx	0,204	0,077	0,29	0,103
	RDFShape	0,311	0,197	0,64	0,118
	Shumlex	0,441	0,380	1	0,161

Los estadísticos descriptivos de los resultados cuantitativos para el segundo caso de prueba se muestran en la Tabla 3. Las comparaciones de las puntuaciones medias muestran la misma relación entre grupos que observamos anteriormente. Sin embargo, en este caso hay diferencias significativas en los tiempos transcurridos entre los tres grupos ($H(2)=6,05$; $p=0,048$; $\eta^2=0,405$). El *post-hoc* de Dunn determinó diferencias significativas en los tiempos transcurridos entre *Shumlex* y *3DShEx* ($p=0,014$). Si bien el 75% de los usuarios de *Shumlex* logran tiempos más bajos que todos los usuarios de *RDFShape*, las diferencias globales no son significativas ($p=0,242$).

En cuanto concierne a la tasa de éxito y la precisión, no hay diferencias significativas entre los grupos en el segundo caso de prueba ($H(2)=1,78$; $p = 0,41$ y $F(2,10)=2,43$; $p=0,137$).

Tabla 3. Estadísticos descriptivos para los resultados del segundo caso de prueba.

Variable	Grupo	Media	s	max	min
Segundos transcurridos	3DShEx	417,8	173,034	644	247
	RDFShape	265,6	106,746	456	204
	Shumlex	186,5	91,799	314	95
Tasa de éxito	3DShEx	0,583	0,096	0,667	0,5
	RDFShape	0,7	0,14	0,833	0,5
	Shumlex	0,708	0,21	1	0,5
Precisión	3DShEx	0,154	0,0757	0,256	0,074
	RDFShape	0,268	0,074	0,357	0,174
	Shumlex	0,476	0,365	1	0,151

Los estadísticos descriptivos de los resultados cualitativos se muestran en la Tabla 39 del Anexo 10.2. Los valores correspondientes a la propensión al error están invertidos para que la escala de valores sea equivalente al resto. En general, las valoraciones de los usuarios son positivas para todas las herramientas: el 71,1% de las respuestas expresan un nivel de aprobación alto o muy alto. Las herramientas obtienen de media valoraciones neutras o positivas para todas las medidas, a excepción de la Escalabilidad, que obtiene valoraciones neutras o negativas de media. El análisis estadístico no mostró diferencias significativas entre los tres grupos para ninguna medida.

6.3 Discusión

Los resultados del primer caso de prueba no muestran ninguna diferencia significativa entre los grupos para ninguna de las métricas. Esto puede ser explicado por la gran variabilidad en todos los grupos -por ejemplo, el tiempo transcurrido para *Shumlex* oscila entre ~1m y ~5m-, así como por la escasa diferencia entre las medias.

No obstante, cabe señalar que sólo un miembro del grupo *Shumlex* obtuvo una puntuación perfecta tanto en el porcentaje de éxito como en la precisión. En general, es probable que la tasa de éxito esté influenciada negativamente por el escaso conocimiento teórico de las *Shape Expressions* que los participantes aseveraron. Esto explicaría que ciertas tareas, pese a su sencillez, parezcan causar una confusión generalizada. En particular, la pregunta 9, que consistía en señalar la referencia que conectaba dos *shapes*, no obtuvo ninguna respuesta correcta por parte de los grupos *3DShEx* y *RDFShape*, mientras que la mayoría de los usuarios de *Shumlex* respondieron correctamente. Dada su singularidad dentro del experimento, esta diferencia concreta en el rendimiento puede deberse a una causa subyacente que desconocemos o al puro azar.

Por lo tanto, sólo podemos afirmar con certeza que en los casos de baja complejidad diagramática no hay evidencia de diferencias en el rendimiento entre las herramientas. Dado que sólo el grupo de *Shumlex* consiguió completar todas las tareas propuestas y tuvo un rendimiento adecuado en las tareas propensas a errores, podría ser necesaria una evaluación adicional con muestras más grandes para evaluar la posible influencia de la herramienta en las tasas de éxito y la precisión.

En cuanto al segundo caso de prueba, los tiempos transcurridos muestran diferencias significativas entre los grupos con un gran tamaño del efecto ($\eta^2=0,405$). Los resultados *post-hoc* sugieren que los usuarios de *3DShEx* necesitan más tiempo para realizar tareas en casos grandes que los usuarios de *Shumlex*. Esto puede explicarse por la combinación de un espacio navegable novedoso con una gran complejidad diagramática que provoca una alta carga cognitiva, superando así los límites de la memoria de trabajo. Esta hipótesis de la alta carga cognitiva se ve apoyada por el hecho de que el único usuario de *3DShEx* con menor capacidad espacial obtuvo el mayor tiempo en el experimento. Como se indica en el apartado 3.2, una menor capacidad espacial puede implicar una mayor carga cognitiva en entornos 3D. El mayor rendimiento de tiempo de los usuarios

de *Shumlex* puede deberse a su mayor parecido con los diagramas de clase UML, con cuya especificación los usuarios afirmaron estar familiarizados.

En cuanto a las tasas de éxito y la precisión, los resultados no muestran diferencias significativas entre los grupos. Al igual que en la primera prueba, sólo un miembro del grupo *Shumlex* obtuvo una puntuación perfecta tanto en la tasa de éxito como en la precisión. Además, la mayoría de los usuarios de *Shumlex* obtienen una precisión superior a la de los otros grupos (véase la Fig. 8). Una vez más, esto puede insinuar una posible influencia de *Shumlex* en las tasas de éxito y precisión también para los casos más grandes. Sea como fuere, la ausencia de pruebas sólidas hace que las afirmaciones sobre la generalización de estos resultados deban restringirse hasta que se realicen más evaluaciones en el futuro.

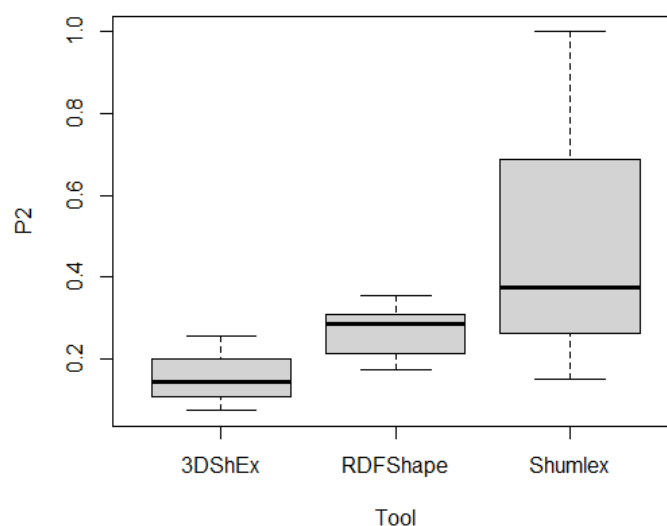


Figura 8. Precisión en el segundo caso según la herramienta.

Por último, los resultados no muestran diferencias significativas entre los grupos en ninguna de las variables de la evaluación subjetiva realizada por los alumnos. Las valoraciones medias son mayoritariamente positivas o neutras en su defecto. La única excepción en el análisis cualitativo es la variable Escalabilidad¹³, en la que las herramientas obtuvieron una puntuación neutra o negativa.

Esto sugiere que los usuarios perciben que las herramientas de visualización son más útiles con esquemas pequeños o medianos. Aunque las herramientas de gestión de la complejidad parecen ser apreciadas, la carga cognitiva extrínseca puede seguir siendo

¹³ “La herramienta es más útil en casos de uso grandes.”

excesiva. Este efecto se ve probablemente exacerbado por su falta de familiaridad con el lenguaje y la notación.

Curiosamente, aunque RDFSshape obtiene una puntuación de 3,8/5 en Gestión de la Complejidad, cuando se les pide su opinión algunos de sus usuarios no transmiten sino insatisfacción en este sentido. *"Como es de esperar, a más grande el caso de uso, más confuso el diagrama"* o *"En grafos muy grandes a veces es complicado ver las flechas que relacionan entidades en las partes centrales"*. A pesar de esas -esperadas- afirmaciones, ninguna de ellas lo califica negativamente.

Ante tales contradicciones, hay varias explicaciones posibles. Puede que no hayan entendido del todo el enunciado a evaluar¹⁴, o puede que hayan temido que una valoración excesivamente severa resulte perjudicial para nuestros intereses (un usuario de *RDFSshape* llega a dar una puntuación perfecta a todas las variables menos una). Tal vez sea simplemente una consecuencia de su falta de experiencia. En cualquier caso, pareciera que los comentarios de los usuarios contienen información de mayor valor para el estudio.

Un análisis más detallado de los comentarios de los usuarios revela una percepción similar de la gestión de la complejidad de *3DShEx*. *"En casos pequeños es muy útil, en casos tan grandes como el segundo es super difícil de tratar"* y *"En el segundo caso era imposible seguir las relaciones y encontrar los textos de las relaciones de cada shape"*. No podemos determinar si realmente utilizaron los mecanismos de gestión de la complejidad. Sin embargo, si ese es el caso, éstos pueden resultar poco intuitivos para los usuarios; lo cierto es que la herramienta no consigue satisfacer esas necesidades. En comparación, hay un único comentario relacionado con *Shumlex*: *"En el segundo caso, [...] puede ser un poco complicado discernir el nombre de la relación. Se puede seleccionar la shape de la cual sale para diferenciar [el nombre] pero estaría bien poder diferenciar el nombre de esta al hacer click sobre ella o haciendo hover"*. Se aprecia la aportación del mecanismo a la vez que se ofrecen soluciones alternativas a esa tarea concreta.

Como síntesis del análisis cualitativo, la percepción de la gestión de la complejidad parece ser mejor para *Shumlex*, mientras que el resto de variables parecen tener un impacto similar en todas las herramientas.

¹⁴ "La herramienta facilita el entendimiento de áreas complejas."

7 Conclusiones y trabajo futuro

Hemos propuesto una notación visual basada en UML para ShEx construida sobre principios ampliamente utilizados y operacionalizados. Además, dicha notación se ha implementado en sendos prototipos 2D y 3D, denominados respectivamente *Shumlex* y *3DShEx*. Los resultados del análisis cualitativo y cuantitativo llevan a las siguientes conclusiones:

Eficiencia de Shumlex. Aunque los usuarios de *Shumlex* obtuvieron en su mayoría mejores resultados que los participantes con otras herramientas, independientemente del caso de prueba, el pequeño tamaño de la muestra implica que esas diferencias no fueron lo suficientemente significativas como para poder generalizar afirmaciones de mayor eficiencia. No obstante, lo anterior, junto con la recepción de los comentarios más positivos de los usuarios, nos hace pensar que esa universalización puede ser posible con más investigación. Es probable que la ausencia de quejas generalizadas sobre la gestión de la complejidad -como ocurre en las otras- sea el resultado de los mecanismos puestos en marcha.

Sobrecarga cognitiva en el entorno 3D. Los usuarios de *3DShEx* fueron significativamente más lentos que los de *Shumlex* en un caso de uso grande. Además, *3DShEx* obtiene las peores calificaciones medias en tasa de éxito, precisión y la mayoría de las variables cualitativas. Aunque las diferencias no son estadísticamente significativas en ellas, consideramos que es probable que nuevas investigaciones puedan servir de base para su confirmación. Por último, a pesar de tener a su disposición un mecanismo de gestión de la complejidad similar al de *Shumlex*, las quejas de los usuarios se dirigen a la escalabilidad. Dado que la carga cognitiva intrínseca es la misma que la de *Shumlex*, concluimos que la interacción con el entorno 3D causa una mayor carga cognitiva extrínseca al usuario. La sobrecarga cognitiva resultante frustra al usuario en detrimento de su comprensión y del uso adecuado de las funcionalidades a su disposición.

Notación visual similar a UML. Al diseñar la notación visual de tal modo que se asemeje a los diagramas de clase UML, esperábamos conseguir una solución intuitiva y transparente sin renunciar a la eficiencia. La evaluación de los usuarios sobre la facilidad de aprendizaje y la transparencia semántica es razonablemente afirmativa de

tal intención. Por otra parte, su eficacia parece depender más bien de la forma en que se presenta la notación visual, dadas las diferencias entre *Shumlex* y *3DShEx*.

Trabajo futuro. Creemos que los esfuerzos futuros deben centrarse en *Shumlex* como la aproximación más prometedora. El análisis de los comentarios de los usuarios sugiere una ampliación de los mecanismos de gestión de la complejidad para dar soporte a tareas más especializadas. Por ejemplo, poder seleccionar una única referencia a una *shape*. Además, la inclusión de un motor de búsqueda podría ayudar a los usuarios a navegar por grandes esquemas.

8 Dirección y gestión del TFM

8.1 Planificación del proyecto

8.1.1 Identificación de interesados

La Tabla 4 expone el listado de interesados o *stakeholders* para este proyecto.

Tabla 4. Lista de interesados

Nombre	Cargo	Departamento
Jorge Álvarez Fidalgo	Investigador	Universidad de Oviedo
José Emilio Labra Gayo	Jefe de proyecto	Universidad de Oviedo
W3C ShEx Community Group	Externo	N/A

8.1.2 OBS, PBS

Los recursos del OBS (*Organization Breakdown Structure*) se muestran en la Tabla 5.

Tabla 5. Recursos del OBS

Nombre	Tipo	Iniciales	Capacidad máxima	Tasa estándar	Acumular
Investigador (Jorge Álvarez Fidalgo)	Trabajo	I	100%	50,00 €/hora	Prorratio
Jefe de proyecto (José Emilio Labra Gayo)	Trabajo	J	20%	60,00 €/hora	Prorratio

El PBS del proyecto se muestra en la Fig. 9.

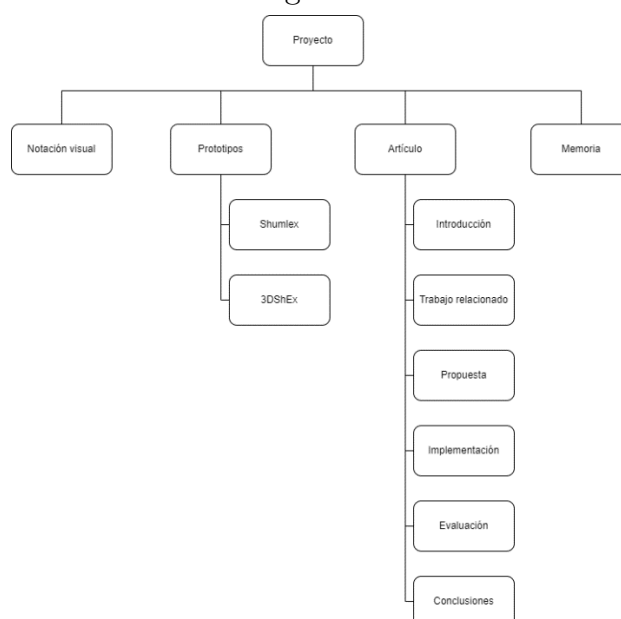


Figura 9. PBS del proyecto.

8.1.3 Planificación inicial. WBS

En la Fig. 10 se muestra la planificación inicial en forma de diagrama de Gantt. Comienza el lunes 24 de enero de 2022 y finaliza el viernes 29 de abril del mismo año. Consta de 55 tareas secuenciales agrupadas en 7 fases. El detalle de las mismas se expone en la Tabla 6.

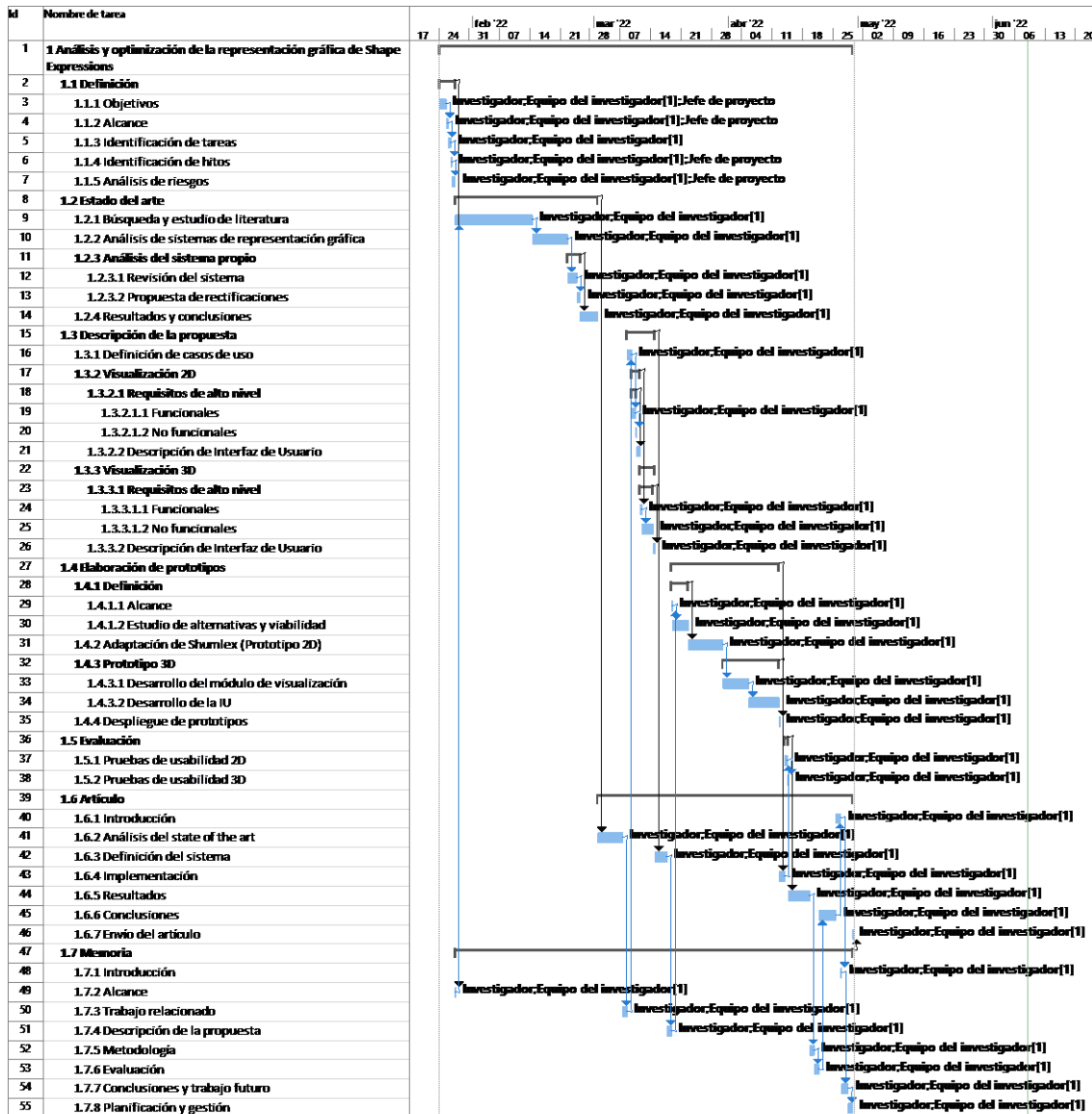


Figura 10. Diagrama de Gantt de la planificación inicial

Tabla 6. Planificación inicial

ID	Tarea	Inicio	Finalización	Duración
1	Análisis y optimización de la representación gráfica de Shape Expressions	lun 24/01/22	vie 29/04/22	556 horas
1.1	Definición	lun 24/01/22	jue 27/01/22	29 horas
1.1.1	Objetivos	lun 24/01/22	mar 25/01/22	10 horas
1.1.2	Alcance	mar 25/01/22	mar 25/01/22	6 horas
1.1.3	Identificación de tareas	mié 26/01/22	mié 26/01/22	5 horas
1.1.4	Identificación de hitos	mié 26/01/22	mié 26/01/22	2 horas
1.1.5	Gestión de riesgos	mié 26/01/22	jue 27/01/22	6 horas
1.2	Estado del arte	jue 27/01/22	mar 01/03/22	180 horas
1.2.1	Búsqueda y estudio de literatura	jue 27/01/22	lun 14/02/22	90 horas
1.2.2	Análisis de sistemas de representación gráfica	lun 14/02/22	mar 22/02/22	50 horas
1.2.3	Análisis del sistema propio	mar 22/02/22	vie 25/02/22	25 horas
1.2.3.1	Revisión del sistema	mar 22/02/22	jue 24/02/22	20 horas
1.2.3.2	Propuesta de rectificaciones	jue 24/02/22	vie 25/02/22	5 horas
1.2.4	Resultados y conclusiones	vie 25/02/22	mar 01/03/22	15 horas
1.3	Descripción de la propuesta	mar 08/03/22	lun 14/03/22	36 horas
1.3.1	Definición de casos de uso	mar 08/03/22	mié 09/03/22	8 horas
1.3.2	Visualización 2D	mié 09/03/22	vie 11/03/22	14 horas

1.3.2.1	Requisitos de alto nivel	mié 09/03/22	jue 10/03/22	8 horas
1.3.2.1.1	Funcionales	mié 09/03/22	jue 10/03/22	6 horas
1.3.2.1.2	No funcionales	jue 10/03/22	jue 10/03/22	2 horas
1.3.2.2	Descripción de Interfaz de Usuario	jue 10/03/22	vie 11/03/22	6 horas
1.3.3	Visualización 3D	vie 11/03/22	lun 14/03/22	14 horas
1.3.3.1	Requisitos de alto nivel	vie 11/03/22	lun 14/03/22	8 horas
1.3.3.1.1	Funcionales	vie 11/03/22	vie 11/03/22	6 horas
1.3.3.1.2	No funcionales	vie 11/03/22	lun 14/03/22	2 horas
1.3.2.2	Descripción de Interfaz de Usuario	lun 14/03/22	lun 14/03/22	6 horas
1.4	Elaboración de prototipos	vie 18/03/22	mar 12/04/22	136 horas
1.4.1	Definición	vie 18/03/22	mar 22/03/22	14 horas
1.4.1.1	Alcance	vie 18/03/22	vie 18/03/22	4 horas
1.4.1.2	Estudio de alternativas y viabilidad	vie 18/03/22	mar 22/03/22	10 horas
1.4.2	Adaptación de Shumlex (Prototipo 2D)	mar 22/03/22	mié 30/03/22	50 horas
1.4.3	Prototipo 3D	mié 30/03/22	mar 12/04/22	70 horas
1.4.3.1	Desarrollo del módulo de visualización	mié 30/03/22	mar 05/04/22	30 horas
1.4.3.2	Desarrollo de la IU	mar 05/04/22	mar 12/04/22	40 horas
1.4.4	Despliegue de prototipos	mar 12/04/22	mar 12/04/22	2 horas
1.5	Evaluación	mié 13/04/22	jue 14/04/22	8 horas
1.5	Pruebas de usabilidad 2D	mié 13/04/22	jue 14/04/22	4 horas
1.5.1	Pruebas de usabilidad 3D	jue 14/04/22	jue 14/04/22	4 horas

1.6	Artículo	mar 01/03/22	vie 29/04/22	345 horas
1.6.1	Introducción	lun 25/04/22	mar 26/04/22	10 horas
1.6.2	Análisis del state of the art	mar 01/03/22	lun 07/03/22	30 horas
1.6.3	Definición del sistema	lun 14/03/22	jue 17/03/22	20 horas
1.6.4	Implementación	mar 12/04/22	mié 13/04/22	10 horas
1.6.5	Resultados	jue 14/04/22	mar 19/04/22	20 horas
1.6.6	Conclusiones	jue 21/04/22	lun 25/04/22	15 horas
1.6.7	Envío del artículo	vie 29/04/22	vie 29/04/22	2 horas
1.7	Memoria	jue 27/01/22	vie 29/04/22	525 horas
1.7.1	Introducción	mar 26/04/22	mar 26/04/22	2 horas
1.7.2	Alcance	jue 27/01/22	jue 27/01/22	2 horas
1.7.3	Trabajo relacionado	lun 07/03/22	mar 08/03/22	10 horas
1.7.4	Descripción de la propuesta	jue 17/03/22	vie 18/03/22	8 horas
1.7.5	Metodología	mar 19/04/22	mié 20/04/22	10 horas
1.7.6	Evaluación	mié 20/04/22	jue 21/04/22	10 horas
1.7.7	Conclusiones y trabajo futuro	Mié 27/04/22	jue 28/04/22	10 horas
1.7.8	Planificación y gestión	jue 28/04/22	jue 29/04/22	10 horas

8.1.4 Riesgos

8.1.4.1. Plan de gestión de riesgos

El Plan de gestión de riesgos se incluye en su totalidad en el anexo 10.3.

8.1.4.2. Identificación de riesgos

En la Tabla 7 se muestran los riesgos identificados para este proyecto.

Tabla 7. Riesgos identificados

ID	Nombre
1	Definición de un nuevo sistema no justificada
2	Los usuarios objetivo consideran que el prototipo no resuelve las carencias y/o problemas de los sistemas precedentes.
3	La información obtenida a lo largo del proyecto resulta insuficiente para elaborar un artículo académico significativo.
4	Falta de literatura en el estudio inicial
5	Falta de voluntarios para participar en las pruebas de usabilidad
6	Los resultados del proyecto no se ajustan a lo esperado
7	Aparición de una nueva herramienta para la visualización de <i>Shape Expressions</i>
8	Incompletitud del prototipo en período lectivo
9	Obtención incompleta de datos durante la evaluación
10	Cambio en las librerías
11	Falta de material para realizar el experimento
12	Fallo de los prototipos durante el experimento
13	Detección de errores en la notación tras implementación de los prototipos
14	La notación visual definida es demasiado compleja para su implementación en el tiempo establecido
15	Imposibilidad de envío a revista

8.1.4.3. Registro de riesgos

Tabla 8. Registro de riesgos

ID	Responsable	Prob.	Impacto				Impacto	0,50 Priorización	Respuesta
			Presup.	Planific.	Alcance	Calidad			
1	Investigador	Media	Muy Bajo	Bajo	Crítico	Bajo	0,45	Estrategia proactiva de mitigación. Llevar a cabo un estudio preliminar tal que el análisis se realice sobre sistemas con cierta constancia de fallo.	
2	Investigador	Media	Medio	Medio	Muy Bajo	Alto	0,28	Estrategia proactiva de mitigación. En la fase de diseño, contactar con los usuarios objetivo para obtener su visión al respecto de los aspectos susceptibles de mejora, y plasmar lo obtenido en los requisitos del sistema.	
3	Investigador	Baja	Muy Bajo	Bajo	Alto	Alto	0,17	Estrategia proactiva de mitigación. Se debe procurar la obtención de datos tangibles en todas las fases del proyecto, sobre los aspectos funcionales y no funcionales de los sistemas a examen, así como el propuesto; tiempos de respuesta, nº de sistemas interoperables, resultados de pruebas de usabilidad... con los cuales sentar las bases para el artículo.	
4	Investigador	Media	Muy Bajo	Bajo	Medio	Bajo	0,15	Estrategia de aceptación. Se trata de un factor de difícil resolución, dado lo específico de la temática y nuestra incapacidad para influir en el volumen de trabajos existentes.	
5	Investigador	Media	Bajo	Medio	Medio	Alto	0,28	Estrategia proactiva de evitación. Contacto con antelación con los miembros de la comunidad de Web Semántica cuyo interés en la materia nos es conocido y cerrar su participación en una fecha de mutuo acuerdo.	
6	Investigador	Alta	Muy Bajo	Bajo	Bajo	Medio	0,21	Estrategia reactiva de mitigación. En el marco de la investigación, es asumible e incluso probable que los resultados obtenidos al final del proyecto no se correspondan con los esperados. Hemos de procurar, no obstante, una correcta justificación de aquellos y explicar	

									razonadamente el motivo de las diferencias acaecidas.
7	Investigador	Baja	Muy Bajo	Bajo	Bajo	Medio	0,09		Estrategia proactiva de mitigación. Actualizar el trabajo relacionado para incluir la nueva herramienta.
8	Investigador	Baja	Bajo	Medio	Medio	Medio	0,09		Estrategia proactiva de evitación. Procurar un margen suficiente de tiempo entre la implementación y la evaluación.
9	Investigador	Media	Muy Bajo	Muy Bajo	Medio	Alto	0,28		Estrategia proactiva de mitigación. Mayor detalle en las variables medidas para disponer de un margen de maniobra más amplio. Exponer claramente los pasos a seguir a los participantes.
10	Investigador	Alta	Bajo	Bajo	Bajo	Bajo	0,11		Estrategia proactiva de evitación. Se elegirá la última versión estable de las librerías, ciñéndose a ellas independientemente de las novedades que pudieran surgir.
11	Investigador	Media	Bajo	Bajo	Medio	Medio	0,15		Estrategia proactiva de mitigación. Pedir a los participantes que aporten sus ordenadores personales el día de la prueba. Como plan de contingencia, asegurarse que los equipos de la clase dispongan de todo el software necesario.
12	Investigador	Media	Bajo	Bajo	Medio	Medio	0,15		Estrategia proactiva de mitigación. Realizar pruebas preliminares para cerciorarse del correcto funcionamiento para las tareas solicitadas.
13	Investigador	Media	Medio	Medio	Medio	Alto	0,28		Estrategia proactiva de mitigación. Dedicar mayores recursos al diseño de una notación visual eficiente, de acuerdo a un marco de referencia tangible.
14	Investigador	Media	Medio	Medio	Medio	Medio	0,15		Estrategia proactiva de mitigación. Realizar una notación visual con pocos símbolos gráficos y, preferiblemente, de uso común, tal que su materialización resulte más sencilla.
15	Investigador	Baja	Medio	Medio	Medio	Medio	0,09		Estrategia proactiva de mitigación. Establecer un margen temporal entre la realización del artículo y su envío, tal que exista cierto margen de maniobra en caso de imprevistos.

8.1.5 Presupuesto inicial

El presupuesto se calcula con los costes por recurso establecidos en la Tabla 5.

8.1.5.1. Partidas de gasto

Tabla 9. Partida de Definición

I 1	I 2	I 3	Descripción	Cantidad	Unidades	Precio	Subtotal (3)	Subtotal (2)	Total
1			Objetivos						1.100,00 €
	1		Investigador	10	horas	50,00 €		500,00 €	
	2		Jefe de proyecto	10	horas	60,00 €		600,00 €	
2			Alcance						660,00 €
	1		Investigador	6	horas	50,00 €		300,00 €	
	2		Jefe de proyecto	6	horas	60,00 €		360,00 €	
3			Identificación de tareas						250,00 €
	1		Investigador	5	horas	50,00 €		250,00 €	
4			Identificación de hitos						220,00 €
	1		Investigador	2	horas	50,00 €		100,00 €	
	2		Jefe de proyecto	2	horas	60,00 €		120,00 €	
5			Gestión de riesgos						660,00 €
	1		Investigador	6	horas	50,00 €		300,00 €	
	2		Jefe de proyecto	6	horas	60,00 €		360,00 €	
TOTAL									2.890,00 €

Tabla 10. Partida de Estado del arte

I 1	I 2	I 3	Descripción	Cantidad	Unidades	Precio	Subtotal (3)	Subtotal (2)	Total
1			Búsqueda y estudio de literatura						4.500,00 €
	1		Investigador	90	horas	50,00 €		4.500,00 €	
2			Análisis de sistemas de representación gráfica						2.500,00 €
	1		Investigador	50	horas	50,00 €		2.500,00 €	
3			Análisis del sistema propio						1.250,00 €
	1		Revisión del sistema					1.000,00 €	
		1	Investigador	20	horas	50,00 €	1.000,00 €		
	2		Propuesta de rectific.					250,00 €	
		1	Investigador	5	horas	50,00 €	250,00 €		
4			Resultados y conclusiones						750,00 €
	1		Investigador	15	horas	50,00 €		750,00 €	
TOTAL									9.000,00 €

Tabla 11. Partida de Descripción del sistema

I	I	I	Descripción	Cantidad	Unidades	Precio	Subtotal (3)	Subtotal (2)	Total
1	2	3							
1			Definición de casos de uso						400,00 €
	1		Investigador	8	horas	50,00 €		400,00 €	
2			Visualización 2D						400,00 €
	1		Requisitos de alto nivel					400,00 €	
		1	Investigador	8	horas	50,00 €	400,00 €		
	2		Descripción de UI					300,00 €	
		1	Investigador	6	horas	50,00 €	300,00 €		
2			Visualización 3D						400,00 €
	1		Requisitos de alto nivel					400,00 €	
		1	Investigador	8	horas	50,00 €	400,00 €		
	2		Descripción de UI					300,00 €	
		1	Investigador	6	horas	50,00 €	300,00 €		
TOTAL									1.200,00 €

Tabla 12. Partida de Elaboración de prototipos

I	I	I	Descripción	Cantidad	Unidades	Precio	Subtotal (3)	Subtotal (2)	Total
1	2	3							
1			Definición						700,00 €
	1		Alcance					200,00 €	
		1	Investigador	4	horas	50,00 €	200,00 €		
	2		Estudio de alternativas y viabilidad					500,00 €	
		1	Investigador	10	horas	50,00 €	500,00 €		
2			Adaptación de Shumlex (Prototipo 2D)						2.500,00 €
	1		Investigador	50	horas	50,00 €		2.500,00 €	
3			Prototipo 3D						3.500,00 €
	1		Desarrollo del módulo de visualización					1.500,00 €	
		1	Investigador	30	horas	50,00 €	1.500,00 €		
	2		Desarrollo de la IU					2.000,00 €	
		1	Investigador	40	horas	50,00 €	2.000,00 €		
4			Despliegue de prototipos						100,00 €
	1		Investigador	2	horas	50,00 €		100,00 €	
TOTAL									6.800,00 €

Tabla 13. Partida de Evaluación

I 1	I 2	I 3	Descripción	Cantidad	Unidades	Precio	Subtotal (3)	Subtotal (2)	Total
1			Pruebas de usabilidad 2D						200,00 €
	1		Investigador	4	horas	50,00 €		200,00 €	
2			Pruebas de usabilidad 3D						200,00 €
	1		Investigador	4	horas	50,00 €		200,00 €	
TOTAL									400,00 €

Tabla 14. Partida de Artículo

I 1	I 2	I 3	Descripción	Cantidad	Unidades	Precio	Subtotal (3)	Subtotal (2)	Total
1			Introducción						500,00 €
	1		Investigador	10	horas	50,00 €		500,00 €	
2			Análisis del state of the art						1.500,00 €
	1		Investigador	30	horas	50,00 €		1.500,00 €	
3			Definición del sistema						1.000,00 €
	1		Investigador	20	horas	50,00 €		1.000,00 €	
3			Implementación						500,00 €
	1		Investigador	10	horas	50,00 €		500,00 €	
4			Resultados						1.000,00 €
	1		Investigador	20	horas	50,00 €		1.000,00 €	
5			Conclusiones						750,00 €
	1		Investigador	15	horas	50,00 €		750,00 €	
6			Envío del artículo						100,00 €
	1		Investigador	2	horas	50,00 €		100,00 €	
TOTAL									5.350,00 €

Tabla 15. Partida de Memoria

I	I	I							
1	2	3	Descripción	Cantidad	Unidades	Precio	Subtotal (3)	Subtotal (2)	Total
1			Introducción						100,00 €
	1		Investigador	2	horas	50,00 €		100,00 €	
2			Alcance						100,00 €
	1		Investigador	2	horas	50,00 €		100,00 €	
2			Trabajo relacionado						500,00 €
	1		Investigador	10	horas	50,00 €		500,00 €	
3			Descripción de la propuesta						400,00 €
	1		Investigador	8	horas	50,00 €		400,00 €	
3			Metodología						500,00 €
	1		Investigador	10	horas	50,00 €		500,00 €	
4			Evaluación						500,00 €
	1		Investigador	10	horas	50,00 €		500,00 €	
5			Conclusiones y trabajo futuro						500,00 €
	1		Investigador	10	horas	50,00 €		500,00 €	
6			Planificación y gestión						400,00 €
	1		Investigador	8	horas	50,00 €		400,00 €	
TOTAL									3.000,00 €

Tabla 16. Partida de Costes indirectos

I	I	I							
1	2	3	Descripción	Cantidad	Unidades	Precio	Subtotal (3)	Subtotal (2)	Total
2			Equipos y licencias						110,00 €
	1		PC					110,00 €	
		1	Amortización	4	meses	27,50 €	110,00 €		
TOTAL									110,00 €

8.1.5.2. Presupuesto de costes

Puesto que se trata de un proyecto de investigación, no se aplica margen de beneficio alguno. La partida 8 (Costes indirectos) suma 110,00 € a repartir entre las 7 partidas restantes que aparecerán en el presupuesto de cliente, cuyos conceptos acumulan 28.640,00 €. Esto supone un incremento del 0,38% sobre los conceptos de las 7 primeras partidas para la obtención del presupuesto de cliente, con un total final de 28.750,00 €.

Tabla 17. Presupuesto de costes

Partidas	Descripción	Coste	Coste Ajustado
1	Definición	2.890,00 €	2.901,10 €
	Objetivos	1.100,00 €	1.104,22 €
	Alcance	660,00 €	662,53 €
	Identificación de tareas	250,00 €	250,96 €
	Identificación de hitos	220,00 €	220,84 €
	Gestión de riesgos	660,00 €	662,53 €
2	Estado del arte	9.000,00 €	9.034,57 €
	Búsqueda y estudio de literatura	4.500,00 €	4.517,28 €
	Análisis de sistemas de representación gráfica	2.500,00 €	2.509,60 €
	Análisis del sistema propio	1.250,00 €	1.254,80 €
	Resultados y conclusiones	750,00 €	752,88 €
3	Descripción de la propuesta	1.200,00 €	1.204,61 €
	Definición de casos de uso	400,00 €	401,54 €
	Visualización 2D	400,00 €	401,54 €
	Visualización 3D	400,00 €	401,54 €
4	Elaboración de prototipos	6.800,00 €	6.826,12 €
	Definición	700,00 €	702,69 €
	Adaptación de Shumlex (Prototipo 2D)	2.500,00 €	2.509,60 €
	Prototipo 3D	3.500,00 €	3.513,44 €
	Despliegue de prototipos	100,00 €	100,38 €
5	Evaluación	400,00 €	401,54 €
	Pruebas de usabilidad 2D	200,00 €	200,77 €
	Pruebas de usabilidad 3D	200,00 €	200,77 €
6	Artículo	5.350,00 €	5.370,55 €
	Introducción	500,00 €	501,92 €
	Análisis del state of the art	1.500,00 €	1.505,76 €
	Definición del sistema	1.000,00 €	1.003,84 €
	Implementación	500,00 €	501,92 €
	Resultados	1.000,00 €	1.003,84 €
	Conclusiones	750,00 €	752,88 €
	Envío del artículo	100,00 €	100,38 €
7	Memoria	3.000,00 €	3.011,52 €
	Introducción	100,00 €	100,38 €
	Alcance	100,00 €	100,38 €
	Trabajo relacionado	500,00 €	501,92 €
	Descripción de la propuesta	500,00 €	501,92 €
	Metodología	500,00 €	501,92 €
	Evaluación	400,00 €	401,54 €
	Conclusiones y trabajo futuro	500,00 €	501,92 €
	Planificación y gestión	400,00 €	401,54 €
8	Costes indirectos	110,00 €	N/A
	Equipos y licencias	110,00 €	N/A
TOTAL		28.750,00 €	28.750,00 €

8.1.5.3. Presupuesto de cliente

Tabla 18. Presupuesto de cliente

Partida	Item	Descripción	Importe	Total
1		Definición		€ 2.901,10
	1	Objetivos	1.104,22 €	
	2	Alcance	662,53 €	
	3	Identificación de tareas	250,96 €	
	4	Identificación de hitos	220,84 €	
	5	Gestión de riesgos	662,53 €	
2		Estado del arte		€ 9.034,57
	1	Búsqueda y estudio de literatura	4.517,28 €	
	2	Análisis de sistemas de representación gráfica	2.509,60 €	
	3	Análisis del sistema propio	1.254,80 €	
	4	Resultados y conclusiones	752,88 €	
3		Descripción de la propuesta		€ 1.204,61
	1	Definición de casos de uso	401,54 €	
	2	Visualización 2D	401,54 €	
	3	Visualización 3D	401,54 €	
4		Elaboración de prototipos		€ 6.826,12
	1	Definición	702,69 €	
	2	Adaptación de Shumlex (Prototipo 2D)	2.509,60 €	
	3	Prototipo 3D	3.513,44 €	
	4	Despliegue de prototipos	100,38 €	
5		Evaluación		€ 401,54
	1	Pruebas de usabilidad 2D	200,77 €	
	2	Pruebas de usabilidad 3D	200,77 €	
6		Artículo		€ 5.370,55
	1	Introducción	501,92 €	
	2	Análisis del state of the art	1.505,76 €	
	3	Definición del sistema	1.003,84 €	
	4	Implementación	501,92 €	
	5	Resultados	1.003,84 €	
	6	Conclusiones	752,88 €	
	7	Envío del artículo	100,38 €	
7		Memoria		€ 3.011,52
	1	Introducción	100,38 €	
	2	Alcance	100,38 €	
	3	Trabajo relacionado	501,92 €	
	4	Descripción de la propuesta	501,92 €	
	5	Metodología	501,92 €	
	6	Evaluación	401,54 €	
	7	Conclusiones y trabajo futuro	501,92 €	
	8	Planificación y gestión	401,54 €	
TOTAL				28.750,00 €

8.2 Ejecución del proyecto

8.2.1 Plan de seguimiento de la Planificación

De cara a realizar el Plan de seguimiento, se definen las siguientes líneas base:

- **Inicial.** Se corresponde con la planificación inicial, ya expuesta.
- **Media.** Se considera que el punto medio del proyecto se encuentra justo tras la elaboración de los prototipos. Se actualiza la planificación conforme a las desviaciones de la línea base original señaladas en las reuniones periódicas.
- **Final.** Al término del proyecto se realiza una última actualización de la planificación conforme a los hechos de la segunda parte del mismo. Se revisa en consecuencia el presupuesto y se realiza el informe final de riesgos.

8.2.2 Bitácora de incidencias del proyecto

En la Tabla 19 se exponen las incidencias que han tenido lugar a lo largo del proyecto.

Tabla 19. Bitácora de incidencias

Incidencia	Fecha
Medición incompleta de los tiempos por tarea	23/05/2022

8.2.3 Riesgos

Se realiza un seguimiento de aquellos riesgos que, si bien no poseen un impacto alto que requiera su priorización ($> 0,5$), sí obtienen un valor medio ($> 0,25$). Este criterio comprende a los riesgos de ID 1, 2, 5, 9 y 13. La Tabla 20 expone tales seguimientos. Las tablas 21 a 25 muestran el detalle de los riesgos.

Tabla 20. Seguimiento de riesgos

ID	Nombre	Seguimiento
1	Definición de un nuevo sistema no justificada	El análisis del estado del arte en función de los principios para el diseño de notaciones visuales cognitivamente eficientes reveló problemas varios que justifican la definición de un nuevo sistema.
2	Los usuarios objetivo consideran que el prototipo no resuelve las carencias y/o problemas de los sistemas precedentes.	La evaluación mediante usuarios indica un mayor grado de satisfacción para uno de los prototipos implementados (Shumlex) así como, en general, mejores resultados. No así para 3DShEx, que obtiene peores tiempos y resultados similares a la herramienta original. Puesto que uno de los prototipos parece resolver en cierto grado los problemas señalados, se considera este riesgo superado.
5	Falta de voluntarios para participar en las pruebas de usabilidad	La totalidad de los estudiantes del primer curso del MIW se prestaron a realizar tales pruebas.
9	Obtención incompleta de datos durante la evaluación	Durante la medición de las pruebas, varios usuarios se equivocaron y midieron la totalidad del caso de prueba, en lugar de cada pregunta por separado. Puesto que la medición de las preguntas era un nivel más de granularidad del necesitado para realizar las comparaciones, esto no afecta a la evaluación.
13	Detección de errores en la notación tras implementación de los prototipos	No se detectaron errores en la notación en la etapa indicada.

Tabla 21. Hoja del riesgo 1

ID: 1	Nombre: Definición de un nuevo sistema no justificada					
Descripción: El análisis de los sistemas antecedentes no arroja suficientes aspectos susceptibles de mejora (carencias, problemas) tangibles como para justificar la definición de un nuevo sistema.						
Categoría(s) de riesgo: Tecnología, Dependencias del proyecto.						
Status: Controlado	Causas del Riesgo: <ul style="list-style-type: none"> No se detectan suficientes aspectos problemáticos en los sistemas de representación gráfica analizados. No se dispone de suficientes sistemas de representación gráfica a examinar. La mayoría de los aspectos críticos son excesivamente ambiciosos o incluso irrealizables. 					
Probabilidad	Impacto				Impacto Total	Respuestas
	Presupuesto	Planificación	Alcance	Calidad		
Media	Muy Bajo	Bajo	Crítico	Bajo	0,45	Estrategia proactiva de mitigación. Llevar a cabo un estudio preliminar tal que el análisis se realice sobre sistemas con cierta constancia de fallo.
Probabilidad revisada Fecha	Impacto				Impacto Total	Respuestas
	Presupuesto	Planificación	Alcance	Calidad		
N/A	N/A	N/A	N/A	N/A	N/A	N/A
Riesgos derivados de éste: Finalización prematura del proyecto.						
Riesgo residual: Que, en un intento de garantizar la existencia de suficientes fallos, se establezca un exceso de sistemas a analizar tal que comprometa el tiempo y coste del proyecto.						
Plan de Contingencia: <ol style="list-style-type: none"> Revisión exhaustiva de los sistemas analizados, particularmente en aquellas áreas que en teoría poseen un menor nivel de madurez. <ol style="list-style-type: none"> Tal revisión no debe superar las 10 horas de trabajo (Estimación: $10h * 50 \text{ €} = 500 \text{ €}$). Extracción a partir de la literatura estudiada de aspectos propuestos o no encontrados en los sistemas existentes que pudieran ser empleados. <ol style="list-style-type: none"> Tal estudio no debe superar las 5 horas de trabajo (Estimación: $5h * 50 \text{ €} = 250 \text{ €}$) 					Presupuesto para contingencias: 750 €	
					Planificación temporal de las contingencias: El plan de contingencia tendrá lugar tras llevar a cabo la fase de análisis si se concluye que los hallazgos son insuficientes como para llevar a cabo la siguiente fase de definición de un nuevo sistema. Puesto que previamente se lleva a cabo un estudio de la literatura existente al respecto de la temática, el paso 2 del plan requerirá un menor esfuerzo al trabajar sobre terreno ya reconocido.	
Comentarios: N/A						

Monitorización:					
Se lleva a cabo una monitorización de las horas de trabajo correspondientes a los factores de riesgo previamente descritos. Para este riesgo se han identificado cuatro indicadores: uno asociado a cada factor descrito y uno adicional asociado al consumo de reserva prevista.					
Los indicadores medidos se valoran del siguiente modo:					
	Valor del riesgo				
INDICADORES	Muy alto	Alto	Medio	Bajo	Muy Bajo
Indicador 1: Características mejorables detectadas en el análisis	0,00	3,00	6,00	9,00	>9,00
Indicador 2: Sistemas de representación gráfica a examinar	0,00	1,00	2,00	3,00	>3,00
Indicador 3: % de características intangibles	>70,00%	50,00%	30,00%	20,00%	10,00%
Indicador 4: Consumo de reserva	>80,00%	60,00%	40,00%	20,00%	20,00%
El riesgo se apreciará como el valor más alto de cualquiera de los indicadores.					
Indicadores:					
Indicador 1: Características mejorables detectadas en el análisis. Total de características susceptibles de mejora o faltas en los sistemas de representación gráfica analizados.			Evaluación: Se evalúa tras finalizar el análisis de sistemas de representación gráfica existentes.		
Indicador 2: Sistemas de representación gráfica a examinar Total de sistemas de representación gráfica a ser analizados.			Evaluación: Se evalúa antes de comenzar el análisis de tales sistemas.		
Indicador 3: % de características intangibles Características cuya solución supera el alcance del proyecto de entre el total definido en el indicador 1.			Evaluación: Se evalúa tras finalizar el análisis de sistemas de representación gráfica existentes.		
Indicador 4: Consumo de reserva Importe (€) consumido de la reserva prevista para contingencias.			Evaluación: Se evalúa siempre que se deba asignar dinero de la reserva por contingencias.		

Tabla 22. Hoja del riesgo 2

ID: 2	Nombre: Los usuarios objetivo consideran que el prototipo no resuelve las carencias y/o problemas de los sistemas precedentes.						
Descripción: Se concluye a partir de la evaluación de usuarios que nuestros prototipos no aportan mejora alguna sobre la visualización ya existente.							
Categoría(s) de riesgo: Prestaciones y fiabilidad, calidad							
Status: Controlado	Causas del Riesgo: <ul style="list-style-type: none"> La notación visual posee carencias que afectan a su eficiencia. El prototipo no está diseñado adecuadamente o tiene errores. 						
Probabilidad	Impacto				Impacto Total	Respuestas	
	Presupuesto	Planificación	Alcance	Calidad			
Media	Medio	Medio	Muy Bajo	Alto	0,28	Estrategia proactiva de mitigación. En la fase de diseño, contactar con los usuarios objetivo para obtener su visión al respecto de los aspectos susceptibles de mejora, y plasmar lo obtenido en los requisitos del sistema.	
Probabilidad revisada Fecha	Impacto				Impacto Total	Respuestas	
	Presupuesto	Planificación	Alcance	Calidad			
N/A	N/A	N/A	N/A	N/A	N/A	N/A	
Riesgos derivados de éste: #6							
Riesgo residual: N/A							
Plan de Contingencia: 1. Estudio de la retroalimentación y subsanación de los problemas detectados. a. Tal revisión no debe superar las 20 horas de trabajo (Estimación: 20h * 50 € = 1000 €). 2. Repetición de la evaluación de usuarios con la versión revisada. b. Tal evaluación no debe superar las 3 horas de trabajo (Estimación: 3h * 50 € = 150 €)					Presupuesto para contingencias: 1150 €		
					Planificación temporal de las contingencias: El plan de contingencia tendrá lugar tras llevar a cabo el análisis de los resultados del experimento si estos no arrojan resultados positivos.		
Monitorización:							
			Valor del riesgo				
INDICADORES			Muy alto	Alto	Medio	Bajo	Muy Bajo
Indicador 1: Mejoras detectadas sobre los sistemas previos			0,00	1,00	2,00	3,00	>4,00
Indicador 2: Consumo de reserva			>80,00%	60,00%	40,00%	20,00%	20,00%
El riesgo se apreciará como el valor más alto de cualquiera de los indicadores.							
Indicadores:							
Indicador 1: Mejoras detectadas sobre los sistemas previos.					Evaluación: Se evalúa tras finalizar el análisis de los resultados del experimento.		
Indicador 2: Consumo de reserva					Evaluación: Se evalúa siempre que se deba asignar dinero de la reserva por contingencias.		

Tabla 23. Hoja del riesgo 5

ID: 5	Nombre: Falta de voluntarios para participar en las pruebas de usabilidad						
Descripción: El número de participantes en la evaluación de usuarios es insuficiente como para poder formar grupos con un mínimo de significación.							
Categoría(s) de riesgo: Recursos							
Status: Controlado	Causas del Riesgo: <ul style="list-style-type: none"> No se consigue un acuerdo con la Universidad para que los alumnos participen. El prototipo no está diseñado adecuadamente o tiene errores. 						
Probabilidad	Impacto				Impacto Total	Respuestas	
	Presupuesto	Planificación	Alcance	Calidad			
Media	Medio	Medio	Muy Bajo	Alto	0,28	Estrategia proactiva de mitigación. En la fase de diseño, contactar con los usuarios objetivo para obtener su visión al respecto de los aspectos susceptibles de mejora, y plasmar lo obtenido en los requisitos del sistema.	
Probabilidad revisada Fecha	Impacto				Impacto Total	Respuestas	
	Presupuesto	Planificación	Alcance	Calidad			
N/A	N/A	N/A	N/A	N/A	N/A	N/A	
Riesgos derivados de éste: #9, puesto que sin suficientes participantes los datos obtenidos serán incompletos.							
Riesgo residual: N/A							
Plan de Contingencia: Contacto con el <i>ShEx Community Group</i> para solicitar más voluntarios a realizar mediante prueba en línea. a. No debe superar las 2 horas de trabajo (Estimación: 2h * 50 € = 100 €).				Presupuesto para contingencias: 100 €			
				Planificación temporal de las contingencias: El plan de contingencia tendrá lugar si a una semana de la evaluación no se han asegurado suficientes usuarios.			
Monitorización:							
			Valor del riesgo				
INDICADORES			Muy alto	Alto	Medio	Bajo	Muy Bajo
Indicador 1: Participantes disponibles			6	9	12	15	>18
Indicador 2: Consumo de reserva			>80,00%	60,00%	40,00%	20,00%	20,00%
El riesgo se apreciará como el valor más alto de cualquiera de los indicadores.							
Indicadores:							
Indicador 1: Participantes disponibles				Evaluación: Se evalúa previa realización del experimento.			
Indicador 2: Consumo de reserva				Evaluación: Se evalúa siempre que se deba asignar dinero de la reserva por contingencias.			

Tabla 24. Hoja del riesgo 9

ID: 9	Nombre: Obtención incompleta de datos durante la evaluación					
Descripción: Durante la medición del experimento se produce una pérdida de datos.						
Categoría(s) de riesgo: Prestaciones y fiabilidad						
Status: Controlado	Causas del Riesgo: <ul style="list-style-type: none"> Error en las herramientas de medición. Incomprensión de los usuarios respecto a sus tareas. 					
Probabilidad	Impacto				Impacto Total	Respuestas
	Presupuesto	Planificación	Alcance	Calidad		
Media	Muy Bajo	Muy Bajo	Medio	Alto	0,28	Estrategia proactiva de mitigación. Mayor detalle en las variables medidas para disponer de un margen de maniobra más amplio. Exponer claramente los pasos a seguir a los participantes.
Probabilidad revisada Fecha	Impacto				Impacto Total	Respuestas
	Presupuesto	Planificación	Alcance	Calidad		
N/A	N/A	N/A	N/A	N/A	N/A	N/A
Riesgos derivados de éste: #3						
Riesgo residual: N/A						
Plan de Contingencia: 1. Repetición de la evaluación de usuarios. a. No debe superar las 3 horas de trabajo (Estimación: 3h * 50 € = 150 €).					Presupuesto para contingencias: 150 €	
					Planificación temporal de las contingencias: El plan de contingencia tendrá lugar si se produce el riesgo y los datos obtenidos se consideran insuficientes para el análisis.	
Comentarios: N/A						
Monitorización:						
			Valor del riesgo			
INDICADORES	Muy alto	Alto	Medio	Bajo	Muy Bajo	
Indicador 1: Pérdida de datos	>80,00%	60,00%	40,00%	20,00%	20,00%	
Indicador 2: Consumo de reserva	>80,00%	60,00%	40,00%	20,00%	20,00%	
El riesgo se apreciará como el valor más alto de cualquiera de los indicadores.						
Indicadores:						
Indicador 1: Pérdida de datos					Evaluación: Se evalúa tras la realización del experimento.	
Indicador 2: Consumo de reserva Importe (€) consumido de la reserva prevista para contingencias.					Evaluación: Se evalúa siempre que se deba asignar dinero de la reserva por contingencias.	

Tabla 25. Hoja del riesgo 13

ID: 13	Nombre: Detección de errores en la notación tras implementación de los prototipos						
Descripción: Una vez finalizados los prototipos, se observan errores o problemas en la notación visual definida que afecta a su eficiencia cognitiva.							
Categoría(s) de riesgo: Prestaciones y fiabilidad, calidad							
Status: Controlado	Causas del Riesgo: <ul style="list-style-type: none"> • Diseño defectuoso. • Aplicación incorrecta de los principios operacionalizados. 						
Probabilidad	Impacto				Impacto Total	Respuestas	
	Presupuesto	Planificación	Alcance	Calidad			
Media	Medio	Medio	Medio	Alto	0,28	Estrategia proactiva de mitigación. Mayor detalle en las variables medidas para disponer de un margen de maniobra más amplio. Exponer claramente los pasos a seguir a los participantes.	
Probabilidad revisada Fecha	Impacto				Impacto Total	Respuestas	
	Presupuesto	Planificación	Alcance	Calidad			
N/A	N/A	N/A	N/A	N/A	N/A	N/A	
Riesgos derivados de éste: #8, #14							
Riesgo residual: N/A							
Plan de Contingencia: Rediseño de la notación y cambios pertinentes en los prototipos. No debe superar las 25 horas de trabajo (Estimación: 25h * 50 € = 1250 €).					Presupuesto para contingencias: 1250 €		
					Planificación temporal de las contingencias: El plan de contingencia tendrá lugar si se detecta el riesgo al menos una semana antes de la evaluación.		
Comentarios: N/A							
Monitorización:							
			Valor del riesgo				
INDICADORES			Muy alto	Alto	Medio	Bajo	Muy Bajo
Indicador 1: Errores detectados			4	3	2	1	0
Indicador 2: Consumo de reserva			>80,00%	60,00%	40,00%	20,00%	20,00%
El riesgo se apreciará como el valor más alto de cualquiera de los indicadores.							
Indicadores:							
Indicador 1: Errores detectados					Evaluación: Se evalúa tras la implementación de los prototipos.		
Indicador 2: Consumo de reserva					Evaluación: Se evalúa siempre que se deba asignar dinero de la reserva por contingencias.		

8.3 Cierre del proyecto

8.3.1 Planificación final

En la planificación final, expuesta en la Tabla 26, tienen lugar una serie de cambios. Si bien la estructura general es similar a la inicial, varias tareas -particularmente relativas a Propuesta e Implementación- se modifican o eliminan para ajustarse a la realidad de su materialización. P.e. surgen tareas relativas al diseño e implementación de mecanismos de gestión de la complejidad, cuya necesidad es evidente tras estudiar el trabajo relacionado. El orden de las mismas y las fechas también se alteran para reflejar los hechos. Las tareas suman un total de 550 horas *trabajadas*.



Figura 11. Diagrama de Gantt de la planificación final

Tabla 26. Planificación final

ID	Tarea	Inicio	Finalización	Duración
1	Análisis y optimización de la representación gráfica de Shape Expressions	lun 24/01/22	dom 05/06/22	760 horas
1.1	Definición	lun 24/01/22	vie 11/03/22	275 horas
1.1.1	Objetivos	lun 24/01/22	lun 24/01/22	5 horas
1.1.2	Alcance	lun 24/01/22	mar 25/01/22	4 horas
1.1.3	Identificación de tareas	mar 08/03/22	mié 09/03/22	5 horas
1.1.4	Identificación de hitos	mié 09/03/22	mié 09/03/22	2 horas
1.1.5	Gestión de riesgos	mié 09/03/22	vie 11/03/22	12 horas
1.2	Estado del arte	vie 11/03/22	vie 01/04/22	120 horas
1.2.1	Diseño de notaciones visuales	vie 11/03/22	mié 23/03/22	30 horas
1.2.2	Visualizaciones en la Web Semántica	mié 23/03/22	vie 01/04/22	30 horas
1.3	Propuesta	jue 07/04/22	lun 18/04/22	55 horas
1.3.1	Diseño de notación visual	jue 07/04/22	vie 15/04/22	25 horas
1.3.2	Mecanismos de gestión de la complejidad	vie 15/04/22	lun 18/04/22	6 horas
1.3.2.1	Shumlex	vie 15/04/22	vie 15/04/22	3 horas
1.3.2.2	3DShEx	vie 15/04/22	lun 18/04/22	3 horas
1.4	Implementación	lun 25/04/22	jue 12/05/22	104 horas
1.4.1	Definición	lun 25/04/22	mar 26/04/22	10 horas
1.4.1.1	Alcance	lun 25/04/22	lun 25/04/22	1 hora
1.4.1.2	Estudio de alternativas y viabilidad	lun 25/04/22	mar 26/04/22	5 horas
1.4.2	Prototipo Shumlex	mar 26/04/22	vie 29/04/22	24 horas
1.4.2.1	Desarrollo del módulo de transformación	mar 26/04/22	mié 27/04/22	8 horas

1.4.2.2	Desarrollo del módulo de visualización	mié 27/04/22	jue 28/04/22	8 horas
1.4.2.3	Gestión de la complejidad	jue 28/04/22	vie 29/04/22	8 horas
1.4.3	Prototipo 3DShEx	lun 02/05/22	mié 11/05/22	62 horas
1.4.3.1	Desarrollo del módulo de transformación	lun 02/05/22	mar 03/05/22	8 horas
1.4.3.2	Desarrollo del módulo de visualización	mié 04/05/22	mar 10/05/22	20 horas
1.4.3.3	Gestión de la complejidad	mar 10/05/22	mié 11/05/22	10 horas
1.4.4	Despliegue de prototipos	jue 12/05/22	jue 12/05/22	2 horas
1.5	Evaluación	lun 23/05/22	lun 23/05/22	3 horas
1.5.1	Evaluación intra-sujetos	lun 23/05/22	lun 23/05/22	3 horas
1.6	Artículo	vie 01/04/22	dom 05/06/22	365 horas
1.6.1	Introducción	mar 31/05/22	mié 01/06/22	10 horas
1.6.2	Ejemplo motivador	vie 01/04/22	vie 01/04/22	2 horas
1.6.3	Trabajo relacionado	lun 04/04/22	jue 07/04/22	15 horas
1.6.4	Propuesta	lun 18/04/22	lun 25/04/22	20 horas
1.6.5	Implementación	jue 12/05/22	lun 16/05/22	10 horas
1.6.6	Evaluación	lun 23/05/22	lun 30/05/22	20 horas
1.6.7	Conclusiones	lun 30/05/22	mar 31/05/22	5 horas
1.6.8	Envío del artículo	dom 05/06/22	dom 05/06/22	2 horas
1.7	Memoria	mar 17/05/22	dom 05/06/22	112 horas
1.7.1	Introducción	jue 02/06/22	jue 02/06/22	2 horas
1.7.2	Ejemplo motivador	mar 17/05/22	mar 17/05/22	2 horas
1.7.3	Trabajo relacionado	mar 17/05/22	mié 18/05/22	4 horas
1.7.4	Propuesta	mié 18/05/22	jue 19/05/22	4 horas
1.7.5	Implementación	jue 19/05/22	vie 20/05/22	4 horas
1.7.6	Evaluación	mié 01/06/22	jue 02/06/22	4 horas

1.7.7	Conclusiones y trabajo futuro	jue 02/06/22	jue 02/06/22	4 horas
1.7.8	Planificación y gestión	vie 03/06/22	dom 05/06/22	15 horas

8.3.2 Informe final de riesgos

Los riesgos definidos al comienzo de este proyecto fueron tratados mediante una estrategia acorde, habitualmente proactiva. El seguimiento de los considerados más relevantes permitió su control a lo largo de la ejecución del proyecto, teniendo lugar únicamente uno de ellos (ID 9). En su caso, la estrategia definida nos permitió mitigar las consecuencias del riesgo. Así pues, el impacto de los riesgos sobre las distintas variables -presupuesto, planificación, alcance y calidad- se determina como muy bajo.

8.3.3 Presupuesto final de costes

A continuación, se muestra el presupuesto correspondiente a la planificación modificada. Se aplican las mismas consideraciones.

Tabla 27. Partida final de Definición

I	I	I	Descripción	Cantidad	Unidades	Precio	Subtotal (3)	Subtotal (2)	Total
1	2	3							
1			Objetivos						550,00 €
	1		Investigador	5	horas	50,00 €		250,00 €	
	2		Jefe de Proyecto	5	horas	60,00 €		300,00 €	
2			Alcance						440,00 €
	1		Investigador	4	horas	50,00 €		200,00 €	
	2		Jefe de Proyecto	4	horas	60,00 €		240,00 €	
3			Identificación de tareas						250,00 €
	1		Investigador	5	horas	50,00 €		250,00 €	
4			Identificación de hitos						220,00 €
	1		Investigador	2	horas	50,00 €		100,00 €	
	2		Jefe de Proyecto	2	horas	60,00 €		120,00 €	
5			Gestión de riesgos						1.320,00 €
	1		Investigador	12	horas	50,00 €		600,00 €	
	2		Jefe de Proyecto	12	horas	60,00 €		720,00 €	
TOTAL									2.780,00 €

Tabla 28. Partida final de Estado del arte

I	I	I	Descripción	Cantidad	Unidades	Precio	Subtotal (3)	Subtotal (2)	Total
1	2	3							
1			Diseño de notaciones visuales						1.500,00 €
	1		Investigador	30	horas	50,00 €		1.500,00 €	

2		Visualizaciones en la Web Semántica						1.500,00 €
1		Investigador	30	horas	50,00 €		1.500,00 €	
TOTAL								3.000,00 €

Tabla 29. Partida final de Propuesta

I 1	I 2	I 3	Descripción	Cantidad	Unidades	Precio	Subtotal (3)	Subtotal (2)	Total
1			Diseño de notación visual						1.250,00 €
	1		Investigador	25	horas	50,00 €		1.250,00 €	
2			Mecanismos de gestión de la complejidad						150,00 €
	1		Shumlex					150,00 €	
		1	Investigador	3	horas	50,00 €	150,00 €		
	2		3DShEx					150,00 €	
		1	Investigador	3	horas	50,00 €	150,00 €		
TOTAL									1.400,00 €

Tabla 30. Partida final de Implementación

I 1	I 2	I 3	Descripción	Cantidad	Unidades	Precio	Subtotal (3)	Subtotal (2)	Total
1			Definición						300,00 €
	1		Alcance					50,00 €	
		1	Investigador	1	horas	50,00 €	50,00 €		
	2		Estudio de alternativas y viabilidad					250,00 €	
		1	Investigador	5	horas	50,00 €	250,00 €		
2			Shumlex						1.200,00 €
	1		Desarrollo del módulo de transformación					400,00 €	
		1	Investigador	8	horas	50,00 €	400,00 €		
	2		Desarrollo del módulo de visualización					400,00 €	
		1	Investigador	8	horas	50,00 €	400,00 €		
	3		Gestión de la complejidad					400,00 €	
		1	Investigador	8	horas	50,00 €	400,00 €		
3			3DShEx						1.900,00 €
	1		Desarrollo del módulo de transformación					400,00 €	
		1	Investigador	8	horas	50,00 €	400,00 €		
	2		Desarrollo del módulo de visualización					1.000,00 €	
		1	Investigador	20	horas	50,00 €	1.000,00 €		
	3		Gestión de la complejidad					500,00 €	
		1	Investigador	10	horas	50,00 €	500,00 €		

4			Despliegue de prototipos						100,00 €
	1		Investigador	2	horas	50,00 €		100,00 €	
TOTAL									3.500,00 €

Tabla 31. Partida final de Evaluación

I	I	I	Descripción	Cantidad	Unidades	Precio	Subtotal (3)	Subtotal (2)	Total
1			Evaluación intra-sujetos						150,00 €
	1		Investigador	3	horas	50,00 €		150,00 €	
TOTAL									150,00 €

Tabla 32. Partida final de Artículo

I	I	I	Descripción	Cantidad	Unidades	Precio	Subtotal (3)	Subtotal (2)	Total
1			Introducción						500,00 €
	1		Investigador	10	horas	50,00 €		500,00 €	
2			Ejemplo motivador						100,00 €
	1		Investigador	2	horas	50,00 €		100,00 €	
3			Trabajo relacionado						750,00 €
	1		Investigador	15	horas	50,00 €		750,00 €	
4			Propuesta						1.000,00 €
	1		Investigador	20	horas	50,00 €		1.000,00 €	
5			Implementación						500,00 €
	1		Investigador	10	horas	50,00 €		500,00 €	
6			Evaluación						1.000,00 €
	1		Investigador	20	horas	50,00 €		1.000,00 €	
7			Conclusiones						250,00 €
	1		Investigador	5	horas	50,00 €		250,00 €	
8			Envío del artículo						100,00 €
	1		Investigador	2	horas	50,00 €		100,00 €	
TOTAL									4.200,00 €

Tabla 33. Partida final de Memoria

I	I	I	Descripción	Cantidad	Unidades	Precio	Subtotal (3)	Subtotal (2)	Total
1			Introducción						100,00 €
	1		Investigador	2	horas	50,00 €		100,00 €	
2			Ejemplo motivador						100,00 €
	1		Investigador	2	horas	50,00 €		100,00 €	
2			Trabajo relacionado						500,00 €

	1		Investigador	10	horas	50,00 €		500,00 €	
3			Propuesta						400,00 €
	1		Investigador	8	horas	50,00 €		400,00 €	
3			Implementación						500,00 €
	1		Investigador	10	horas	50,00 €		500,00 €	
4			Evaluación						500,00 €
	1		Investigador	10	horas	50,00 €		500,00 €	
5			Conclusiones y trabajo futuro						500,00 €
	1		Investigador	10	horas	50,00 €		500,00 €	
6			Planificación y gestión						400,00 €
	1		Investigador	8	horas	50,00 €		400,00 €	
TOTAL									3.000,00 €

Tabla 34. Partida final de Costes indirectos

I	I	I							
1	2	3	Descripción	Cantidad	Unidades	Precio	Subtotal (3)	Subtotal (2)	Total
2			Equipos y licencias						110,00 €
	1		PC					110,00 €	
		1	Amortización	4	meses	27,50 €	110,00 €		
TOTAL									110,00 €

Puesto que se trata de un proyecto de investigación, no se aplica margen de beneficio alguno. La partida 8 (Costes indirectos) suma 110,00 € a repartir entre las 7 partidas restantes que aparecerán en el presupuesto de cliente, cuyos conceptos acumulan 18.030,00 €. Esto supone un incremento del 0,61% sobre los conceptos de las 7 primeras partidas para un total final de 18.140,00 €.

Tabla 35. Presupuesto de costes final

Partidas	Descripción	Coste	Coste Ajustado
1	Definición	2.780,00 €	2.796,96 €
	Objetivos	550,00 €	553,36 €
	Alcance	440,00 €	442,68 €
	Identificación de tareas	250,00 €	251,53 €
	Identificación de hitos	220,00 €	221,34 €
	Gestión de riesgos	1.320,00 €	1.328,05 €
2	Estado del arte	3.000,00 €	3.018,30 €
	Diseño de notaciones visuales	1.500,00 €	1.509,15 €
	Visualizaciones en la Web Semántica	1.500,00 €	1.509,15 €
3	Propuesta	1.400,00 €	1.408,54 €
	Diseño de notación visual	1.250,00 €	1.257,63 €
	Mecanismos de gestión de la complejidad	150,00 €	150,92 €
4	Implementación	3.500,00 €	3.521,35 €
	Definición	300,00 €	301,83 €
	Shumlex	1.200,00 €	1.207,32 €
	3DShEx	1.900,00 €	1.911,59 €
	Despliegue de prototipos	100,00 €	100,61 €
5	Evaluación	150,00 €	150,92 €
	Evaluación intra-sujetos	150,00 €	150,92 €
6	Artículo	4.200,00 €	4.225,62 €
	Introducción	500,00 €	503,05 €
	Ejemplo motivador	100,00 €	100,61 €
	Trabajo relacionado	750,00 €	754,58 €
	Propuesta	1.000,00 €	1.006,10 €
	Implementación	500,00 €	503,05 €
	Evaluación	1.000,00 €	1.006,10 €
	Conclusiones	250,00 €	251,53 €
	Envío del artículo	100,00 €	100,61 €
7	Memoria	3.000,00 €	3.018,30 €
	Introducción	100,00 €	100,61 €
	Ejemplo motivador	100,00 €	100,61 €
	Trabajo relacionado	500,00 €	503,05 €
	Propuesta	500,00 €	503,05 €
	Implementación	500,00 €	503,05 €
	Evaluación	400,00 €	402,44 €
	Conclusiones y trabajo futuro	500,00 €	503,05 €
	Planificación y gestión	400,00 €	402,44 €
8	Costes indirectos	110,00 €	N/A
	Equipos y licencias	110,00 €	N/A
TOTAL		18.140,00 €	18.140,00 €

8.3.4 Presupuesto final de cliente

Tabla 36. Presupuesto de cliente final

Partida	Item	Descripción	Importe	Total
1		Definición		€ 2.796,96
	1	Objetivos	553,36 €	
	2	Alcance	442,68 €	
	3	Identificación de tareas	251,53 €	
	4	Identificación de hitos	221,34 €	
	5	Gestión de riesgos	1.328,05 €	
2		Estado del arte		€ 3.018,30
	1	Diseño de notaciones visuales	1.509,15 €	
	2	Visualizaciones en la Web Semántica	1.509,15 €	
3		Propuesta		€ 1.408,54
	1	Diseño de notación visual	1.257,63 €	
	2	Mecanismos de gestión de la complejidad	150,92 €	
4		Implementación		€ 3.521,35
	1	Definición	301,83 €	
	2	Shumlex	1.207,32 €	
	3	3DShEx	1.911,59 €	
	4	Despliegue de prototipos	100,61 €	
5		Evaluación		€ 150,92
	1	Evaluación intra-sujetos	150,92 €	
6		Artículo		€ 4.225,62
	1	Introducción	503,05 €	
	2	Ejemplo motivador	100,61 €	
	3	Trabajo relacionado	754,58 €	
	4	Propuesta	1.006,10 €	
	5	Implementación	503,05 €	
	6	Evaluación	1.006,10 €	
	7	Conclusiones	251,53 €	
	8	Envío del artículo	100,61 €	
7		Memoria		€ 3.018,30
	1	Introducción	100,61 €	
	2	Ejemplo motivador	100,61 €	
	3	Trabajo relacionado	503,05 €	
	4	Propuesta	503,05 €	
	5	Implementación	503,05 €	
	6	Evaluación	402,44 €	
	7	Conclusiones y trabajo futuro	503,05 €	
	8	Planificación y gestión	402,44 €	
1		Definición		€ 2.796,96
TOTAL				18.140,00 €

8.3.5 Informe de lecciones aprendidas

La evaluación con usuarios es un proceso complejo e impredecible, cuya realización debe ser cuidadosamente planificada. Pese a detallar paso por paso las tareas a realizar, se produjo un error recurrente en la medición de las tareas. Dicho riesgo estaba previsto y dado el mayor nivel de granularidad planteado en las mediciones, esto no impidió el análisis y la extracción de conclusiones del experimento.

No obstante, de cara al futuro debemos cerciorarnos de que los participantes comprenden a la perfección sus labores. La aportación de guías escritas es insuficiente, puesto que tienden a ignorarlas. El responsable debe incidir verbalmente en los aspectos clave.

Por otra parte, esta experiencia revela al autor la dificultad de planificar proyectos de investigación con un gran nivel de detalle, dada su naturaleza errática. Es posible esbozar las fases generales, pero las descripciones específicas fluctúan según aumenta el conocimiento al respecto, particularmente en campos para los que se posee un escaso conocimiento.

9 Referencias

- [1] T. Berners-Lee y M. Fischetti, *Weaving the web: The original design and Ultimate destiny of the World Wide Web by its inventor*, HarperBusiness, 2011.
- [2] E. Prud'hommeaux, J. E. Labra Gayo y H. Solbrig, «Shape expressions: an RDF validation and transformation language,» *Proceedings of the 10th International Conference on Semantic Systems - SEM '14*, 2014.
- [3] C. Ware, «1. Foundations for an Applied Science of Data Visualization,» de *Information visualization: Perception for design*, Morgan Kaufmann, an imprint of Elsevier, 2021.
- [4] J. E. Labra Gayo, D. Fernández Álvarez y H. García González, «RDFShape: An RDF Playground Based on Shapes,» *ISWC 2018 Posters & Demonstrations*, 2018.
- [5] S. Burgstaller-Muehlbacher, A. Waagmeester, E. Mitraka, J. Turner, T. Putman, J. Leong, C. Naik, P. Pavlidis, L. Schriml, B. M. Good y et al., «Wikidata as a semantic framework for the Gene Wiki Initiative,» *Database*, vol. 2016, 2016.
- [6] J. E. Labra Gayo, E. Prudhommeaux, I. Boneva y D. Kontokostas, *Validating RDF data*, Morgan & Claypool Publishers, 2018.
- [7] P. Chandler y J. Sweller, «Cognitive load theory and the format of instruction,» *Cognition and Instruction*, vol. 8, p. 293–332, 1991.
- [8] J. Sweller, «Cognitive load theory, learning difficulty, and instructional design,» *Learning and Instruction*, vol. 4, p. 295–312, 1994.
- [9] D. Moody, «The “physics” of notations: Toward a scientific basis for constructing visual notations in software engineering,» *IEEE Transactions on Software Engineering*, vol. 35, p. 756–779, 2009.
- [10] D. van der Linden y I. Hadar, «A systematic literature review of applications of the physics of notations,» *IEEE Transactions on Software Engineering*, vol. 45, p. 736–759, 2019.
- [11] D. van der Linden, A. Zamansky y I. Hadar, «How cognitively effective is a visual notation? on the inherent difficulty of operationalizing the physics of notations,» *Enterprise, Business-Process and Information Systems Modeling*, p. 448–462, 2016.

- [12] D. van der Linden y I. Hadar, «User involvement in applications of the pon,» *Lecture Notes in Business Information Processing*, p. 109–115, 2016.
- [13] H. Störrle y A. Fish, «Towards an operationalization of the “physics of notations” for the analysis of visual languages,» *Lecture Notes in Computer Science*, p. 104–120, 2013.
- [14] D. Van Der Linden, A. Zamansky y I. Hadar, «A framework for improving the verifiability of visual notation design grounded in the physics of notations,» *2017 IEEE 25th International Requirements Engineering Conference (RE)*, 2017.
- [15] T. Huk, «Who benefits from learning with 3D models? the case of Spatial ability,» *Journal of Computer Assisted Learning*, vol. 22, p. 392–404, 2006.
- [16] T. N. Höffler, «Spatial ability: Its influence on learning with visualizations—a meta-analytic review,» *Educational Psychology Review*, vol. 22, p. 245–269, 2010.
- [17] N. Arndt, A. Valdestilhas, G. Publio, A. Cimmino, K. Höffner y T. Riechert, «A Visual SHACL Shapes Editor Based On OntoPad,» de *Joint Proceedings of the Semantics co-located events: Poster&Demo track and Workshop on Ontology-Driven Conceptual Modelling of Digital Twins co-located with Semantics 2021*, Amsterdam, 2021.
- [18] S. Lieber, De , B. Meester, P. Heyvaert, F. Brückmann, R. Wambacq, E. Mannens, R. Verborgh y A. Dimou, «Visual notations for viewing RDF constraints with unshacled,» *Semantic Web*, p. 1–36, 2021.
- [19] S. Lohmann, S. Negru, F. Haag y T. Ertl, «Visualizing ontologies with Vowl,» *Semantic Web*, vol. 7, p. 399–419, 2016.
- [20] B. Le Grand y M. Soto, «Information management - Topic Maps visualization,» 2000.
- [21] V. Geroimenko, B. Le Grand y M. Soto, «Topic Maps, RDF Graphs, and Ontologies Visualization,» de *Visualizing the semantic web: XML-based internet and information visualization*, Springer, 2004, p. 66–69.
- [22] P. McIntosh, «X3D-UML: User-Centred Design, Implementation and Evaluation of 3D UMLUsing X3D,» *RMIT University*, 2009.
- [23] L. Bartram, R. Ovans, J. Dill, M. Dyck, A. Ho y W. S. Havens, «Contextual Assistance in User Interfaces to Complex, Time-Critical Systems: The Intelligent Zoom,» 1994.
- [24] J. Bertin, *Graphics and graphic information-processing*, W. de Gruyter, 1981.

- [25] G. A. Miller, «The magical number seven, plus or minus two: Some limits on our capacity for processing information.,» *Psychological Review*, vol. 63, p. 81–97, 1956.
- [26] H. García-González, I. Boneva, S. Staworko, J. E. Labra-Gayo y J. M. Cueva Lovelle, «ShExML: Improving the usability of heterogeneous data mapping languages for first-time users,» *PeerJ Computer Science*, vol. 6, 2020.
- [27] J. W. Creswell y J. D. Creswell, *Research design: Qualitative, quantitative, and mixed methods approaches*, SAGE Publications, Inc, 2018.

10 Anexos

10.1 Métrica de similitud

Dada una notación N con símbolos gráficos G_N , el marco de operacionalización de Störrle et al. propone 4 criterios para determinar la discriminabilidad perceptiva: distancia visual (VD), codificación redundante (RC), *pop-out* perceptual (PPO) y diferenciación textual (TD). Estos criterios se normalizan en un intervalo de $[0, 1]$, de tal forma que 0 denote una discriminabilidad nula y 1 el cumplimiento de todos los criterios. Por último, se calcula la discriminabilidad media de la notación.

10.1.1 Distancia visual

Tanto la función de diferencia de variables visuales $vvd(a,b)$ como la función de distancia visual $vd(g,h)$ se utilizan tal cual. El valor máximo de $vd(g,h)$ es 1 y hay G_N^2 combinaciones posibles de g,h . Por lo tanto, la función de distancia visual media $VD(N)$ utiliza G_N^2 como denominador para que 1 sea el valor más alto posible. Esto no tiene en cuenta que siempre que g es igual a h , $vd(g,h) = 0$, por lo que el valor máximo de la suma es $G_N^2 - G_N$. Así, modificamos el denominador y eliminamos la sustracción de la unidad para que los valores se normalicen en el rango especificado.

$$VD(N) := \frac{1}{|G_N^2| - |G_N|} \sum_{g,h \in G_N} vvd(g,h)$$

Por ejemplo, la distancia visual entre los símbolos gráficos *caja* (c) y *flecha dirigida* (fd) es la siguiente. Como se sugiere, los pesos w son 7 para la forma y 1 para el resto de variables visuales. Las variables visuales no utilizadas tienen $vvd = 0$, por lo que sólo podemos calcular la forma, el brillo y la textura. En este caso particular, las formas están en diferentes grupos principales: líneas y regiones. Por tanto, $vvd(v_{sh}(c), v_{sh}(fd)) = 1$. Lo mismo ocurre con el brillo, dado que el color de sus zonas principales es completamente opuesto (blanco y negro). Sin embargo, tienen la misma textura sólida, por lo que $vvd(v_{tx}(c), v_{tx}(fd)) = 0$.

$$vd(c, fd) := \frac{1}{||w||} \sum_{i=1}^d w_i \cdot vvd(v_i(c), v_i(fd)) := \frac{1}{14} (7 \cdot 1 + 1 \cdot 1 + 1 \cdot 0) := 0,57$$

Este mismo procedimiento se repite para todas las combinaciones de símbolos gráficos. La única novedad es la comparación entre formas del mismo grupo básico (es decir, flechas y líneas) para las que $vvd = 0,5$. Sus resultados se muestran en la Tabla 37.

Tabla 37. Distancia visual para los pares de símbolos gráficos.

	Caja	Flecha dirigida	Flecha diamante	Línea entrecortada
Caja	0	0,57	0,57	0,64
Flecha dirigida	0,57	0	0,32	0,32
Flecha diamante	0,57	0,32	0	0,39
Línea entrecortada	0,64	0,32	0,39	0

Con todo lo anterior, podemos finalmente calcular VD para nuestra notación N.

$$VD(N) := \frac{1}{16 - 4} (0,57 \cdot 4 + 0,32 \cdot 4 + 0,64 \cdot 2 + 0,39 \cdot 2) := 0,47$$

10.1.2 Codificación redundante

Los cambios anteriores en el denominador se aplican también a la función RC(N). La función $vr(g,h)$ se utiliza tal cual. Los resultados se muestran en la Tabla 38.

Tabla 38. Codificación redundante para los pares de símbolos gráficos.

	Caja	Flecha dirigida	Flecha diamante	Línea entrecortada
Caja	0	0,25	0,25	0,38
Flecha dirigida	0,25	0	0,25	0,25
Flecha diamante	0,25	0,25	0	0,38
Línea entrecortada	0,38	0,25	0,38	0

$$RC(N) := \frac{1}{16 - 4} (0,25 \cdot 8 + 0,38 \cdot 4) := 0,29$$

10.1.3 Pop-out perceptual

Cada símbolo gráfico tiene al menos un valor único en una variable visual. Únicamente teniendo en cuenta la variable *forma* ya se cumple este criterio. Una vez más, eliminamos la substracción para que el mejor valor posible sea 1. Por lo tanto, la función es la siguiente:

$$PPO(N) := \frac{1}{4} (1 \cdot 4) := 1$$

10.1.4 Diferenciación textual

Como se indica en la Sección 4.1.1, la PoN no considera las construcciones visuales que hacen uso de la diferenciación textual para transmitir significados distintos como símbolos gráficos diferentes (sobrecarga de símbolos). Por lo tanto, se ha modificado este criterio para que no mida la proporción de símbolos gráficos que diferencian únicamente mediante medios textuales. En su lugar, se evalúa la proporción de símbolos gráficos que transmiten varios conceptos mediante la diferenciación textual.

$$TD(N) := 1 - \frac{2}{4} := 0,5$$

10.2 Resultados del análisis cualitativo

Tabla 39. Estadísticos descriptivos de los resultados del análisis cualitativo

Variable	Grupo	Media	s	max	min
Satisfacción general	3DShEx	3	0,82	4	2
	RDFShape	4,4	0,55	5	4
	Shumlex	4,25	0,96	5	3
Facilidad de uso	3DShEx	3,75	1,5	5	2
	RDFShape	4,2	0,45	5	4
	Shumlex	4,5	1	5	3
Aprendizaje	3DShEx	3,75	0,5	4	3
	RDFShape	4	1,22	5	2
	Shumlex	4,5	0,58	5	4
Transparencia semántica	3DShEx	3,75	0,5	4	3
	RDFShape	3,8	0,45	4	3
	Shumlex	4,25	0,5	5	4
Utilidad	3DShEx	3,25	0,96	4	2
	RDFShape	4,2	0,45	5	4
	Shumlex	4	1,15	5	3
Propensión al error	3DShEx	3,25	0,96	4	2
	RDFShape	4,2	0,45	5	4
	Shumlex	4	1,15	5	3
Gestión de la complejidad	3DShEx	3,25	0,96	4	2
	RDFShape	3,6	1,14	5	2
	Shumlex	3,5	1,73	5	2
Escalabilidad	3DShEx	3,5	1,29	5	2
	RDFShape	3,8	0,84	5	3
	Shumlex	3,25	0,96	4	2
Referencias	3DShEx	1,25	0,5	2	1
	RDFShape	3	1,22	5	2
	Shumlex	2,5	1,29	4	1
Restricciones	3DShEx	3,75	0,96	5	3
	RDFShape	4	0,71	5	3
	Shumlex	3,75	0,96	5	3
Global	3DShEx	3	0	3	3
	RDFShape	3,6	1,14	5	2
	Shumlex	3,75	0,96	5	3

10.3 Plan de gestión de riesgos

10.3.1 Metodología

La metodología a emplear durante la elaboración del presente Plan de Gestión de Riesgos se fundamenta en la descrita por Boehm, conformada por las siguientes etapas:

1. Evaluación de los riesgos.

- a. *Identificación de los riesgos.* Proceso de reconocimiento de cualesquiera adversidades que pudieran tener lugar durante el proyecto.
- b. *Análisis de los riesgos.* Definición de la probabilidad del riesgo, así como de su impacto sobre presupuesto, planificación, alcance y calidad. Asignación del responsable pertinente. Identificación de indicadores del riesgo.
- c. *Priorización de riesgos.* Distinguir aquellos riesgos que suponen una auténtica amenaza para el proyecto para proporcionarles una mayor atención en función de su probabilidad e impacto.

2. Control de los riesgos.

- a. *Planificación de los riesgos.* Estudio sobre el modo de evitación o mitigación del impacto -en el caso de los riesgos negativos- o de su explotación -riesgos positivos-.
- b. *Solución a los riesgos.* Definición de planes de contingencia que establezcan los procedimientos en caso de que el riesgo se produzca.
- c. *Monitorización de los riesgos.* Actualización y definición de planes para controlar los riesgos previamente identificados mediante la evaluación de indicadores.

10.3.2 Herramientas y tecnologías

Se utilizarán las siguientes técnicas para llevar a cabo la gestión de riesgos:

- **Tormenta de ideas.** Se empleará para realizar la identificación inicial de la lista principal de los riesgos. Consiste en una discusión sobre los principales objetivos del proyecto y sus posibles devenires, a partir de los cuales extraer los riesgos más obvios.
- **Evaluaciones periódicas.** Examen regular de los indicadores descritos, según el plan establecido en la hoja de riesgo. Sus resultados se discutirán en las reuniones de riesgos.

- **Reuniones.** Con cierta frecuencia, las reuniones periódicas deben tratar los riesgos del proyecto para tomar decisiones sobre aquellos y sus impactos.
- **Checklists.** Evaluación de un listado de cuestiones generales de frecuente ocurrencia para detectar su presencia en el proyecto y tratarlos adecuadamente.
- **Diagramas causa-efecto.** Permiten ilustrar las relaciones entre un riesgo (efecto) y sus factores de riesgo (causas), para facilitar la comprensión de las mismas.
- **Técnica Delphi.** Los expertos en riesgos aportan ideas anónimamente para construir un consenso objetivo.

10.3.3 Roles y responsabilidades

- **Jefe de proyecto.** Se encarga de dirigir y realizar el seguimiento de los riesgos mediante los mecanismos definidos, coordinándose para ello con los responsables y *stakeholders*.
- **Responsable de riesgos.** Se asigna un responsable a cada riesgo, cuya labor consiste en coordinar o ejecutar las respuestas definidas para reducir el impacto de aquel.

10.3.4 Presupuesto

El presupuesto asignado a la Gestión de Riesgos se muestra en la Tabla 40.

Tabla 40. Presupuesto asignado a la Gestión de Riesgos

Item	Concepto	Asignación (€)
1	Identificación de riesgos.	136,66
2	Análisis y priorización de riesgos.	250,00
3	Planificación de riesgos.	136,66
4	Definición de planes de contingencia.	136,66
5	Actualización y monitorización de riesgos.	100,00
TOTAL (€)		760,00

10.3.5 Calendario

La primera parte de la gestión de riesgos está planteada para su realización en el mes de enero, como se expone en la planificación. Esto comprende la identificación, análisis y priorización, planificación y definición de planes de contingencia. En cuanto respecta

a la actualización y monitorización de riesgos, el jefe de proyecto se reunirá semanalmente con los responsables del riesgo para su evaluación y actualización.

10.3.6 Categorías de riesgo

De cara a llevar a cabo la identificación de riesgos y conocer la estructura de aquellos, los riesgos se categorizan acorde a la distinción clásica de tipos de riesgos principales: técnicos, organizacionales, relativos a la gestión de proyectos y externos. Un riesgo puede pertenecer a una o más categorías. La estructura completa se presenta a continuación:

1. Técnico.
 - a. Requisitos.
 - b. Tecnología.
 - c. Prestaciones y fiabilidad.
 - d. Calidad.
 - e. Complejidad.
2. Organizacional.
 - a. Dependencias del proyecto.
 - b. Recursos.
 - c. Financiación.
 - d. Personal.
 - e. Toma de decisiones.
3. Gestión del proyecto.
 - a. Estimación.
 - b. Planificación.
 - c. Control.
 - d. Comunicación.
 - e. Cultural.
4. Externo.
 - a. Proveedores.
 - b. Reguladores.
 - c. Usuario.
 - d. Tiempo.
 - e. Referencias.

10.3.7 Definiciones de probabilidad

Tabla 41. Definiciones de probabilidad

Nombre	% equivalente	Descripción
Muy baja	(0% - 20%]	La probabilidad de que el riesgo tenga lugar es muy poco probable.
Baja	(20% - 40%]	La probabilidad de que el riesgo se produzca es baja.
Media	(40% - 60%]	Existe cierta probabilidad de que el riesgo llegue a ocurrir.
Alta	(60% - 80%]	Hay altas probabilidades de que acaezca la situación que describe el riesgo.
Muy alta	(80% - 100%)	El proyecto sufrirá las consecuencias del riesgo con una alta probabilidad.

10.3.8 Definiciones de impacto por objetivos

Los objetivos seleccionados son los básicos de cualquier proyecto, plasmados en el triángulo de la gestión de proyectos: el alcance, coste y tiempo del proyecto, cuyas relaciones determinarán la calidad del mismo. Tales son los mínimos que hemos de contemplar para la correcta consecución del proyecto. Se considera el impacto tanto para riesgos positivos (amenazas) como negativos (oportunidades). Se emplean las escalas habituales.

Tabla 42. Impacto sobre los objetivos principales

Objetivos de proyecto	Tipo de impacto	Escalas relativas o numéricas				
		Muy bajo	Bajo	Moderado	Alto	Crítico
Alcance	Negativo	El alcance se decrementa de modo apenas perceptible.	El alcance se decrementa en aspectos poco relevantes.	El alcance se decrementa en áreas importantes.	El alcance se decrementa de tal modo que impide la consecución de requisitos de alto nivel.	El alcance se decrementa de tal modo que el proyecto debe concluirse prematuramente.
	Positivo	El alcance se incrementa de modo apenas perceptible	El alcance se incrementa en aspectos poco relevantes.	El alcance se incrementa en áreas importantes.	El alcance se incrementa de tal modo que ayuda a la consecución de requisitos de alto nivel.	El alcance se incrementa de tal modo que se aumenta significativamente la funcionalidad.
Calidad	Negativo	La calidad se decrementa en cuestiones menores.	La calidad se decrementa en navegadores desactualizados.	La calidad se decrementa de tal modo que se requiere la aprobación de instancias superiores.	La calidad se decrementa de tal modo que ha de renunciarse a la presentación del proyecto ante tribunal.	La calidad se decrementa de tal modo que el producto resulta fútil para los intereses de los stakeholders.

	Positivo	Aumento de la calidad de la solución planteada en temas menores.	Aumento de la calidad en navegadores de versiones recientes.	Aumento de la calidad perceptible por todos los usuarios.	Aumento significativo de la calidad.	Gran aumento de la calidad.
Coste	Negativo	Aumento poco notable en el coste.	Aumento en el coste menor al 5%.	Aumento en el coste de entre el 5% y el 10%.	Aumento en el coste de entre el 10% y el 20%.	Aumento en el coste superior al 20%.
	Positivo	Decremento poco notable en el coste.	Decremento en el coste menor al 5%.	Decremento en el coste de entre el 5% y el 10%.	Decremento en el coste de entre el 10% y el 20%.	Decremento en el coste superior al 20%.
Tiempo	Negativo	Aumento en el tiempo de menos de 10 horas.	Aumento de menos de un 5% de lo planificado.	Aumento en el tiempo de entre el 5% y el 10%.	Aumento en el tiempo de entre el 10% y el 20%.	Aumento en el tiempo superior al 20%.
	Positivo	Decremento en el tiempo de menos de 10 horas.	Decremento de menos de un 5% de lo planificado.	Decremento en el tiempo de entre el 5% y el 10%.	Decremento en el tiempo de entre el 10% y el 20%.	Decremento en el tiempo superior al 20%.

10.3.9 Matriz de probabilidad e impacto

Tabla 43. Matriz de probabilidad e impacto

Probabilidad	Muy Alta	0,90	0,05	0,14	0,27	0,50	0,81
	Alta	0,70	0,04	0,11	0,21	0,39	0,63
	Media	0,50	0,03	0,08	0,15	0,28	0,45
	Baja	0,30	0,02	0,05	0,09	0,17	0,27
	Muy Baja	0,10	0,01	0,02	0,03	0,06	0,09
			0,05	0,15	0,30	0,55	0,90
			Muy Bajo	Bajo	Medio	Alto	Crítico
			Impacto				

Los valores son aplicables tanto a impactos positivos como negativos. El impacto global (*score*) se calculará como el mayor de los valores de impacto. Se define el umbral de riesgo en 0,5, sobrepasado lo cual se deben priorizar los riesgos.

10.3.10 Planes de contingencia

- **Presupuesto.** Siempre y cuando sea absolutamente necesario el modificar el presupuesto definido y acordado para el proyecto, el jefe de proyecto ha de examinar y aprobar tales modificaciones, previa condición de que no supongan un exceso superior al 5% del coste original del proyecto.
- **Planificación.** Es posible que se produzcan retrasos en la ejecución de tareas en cuanto el cumplimiento de los siguientes hitos no sea vea afectado:

- El envío del artículo elaborado a una revista JCR. Fecha límite el 6 de junio de 2022.
- La entrega del proyecto final. Fecha límite el 12 de junio de 2022.

10.3.11 Formatos de la documentación

El contenido de este Plan de Gestión de Riesgos está basado en la guía proporcionada por el PMBOK en su sexta versión. Asimismo, para la gestión de la documentación a elaborar durante la ejecución del proyecto se tomarán como referencia las siguientes normas:

- UNE-ISO 31000:2018. Gestión del riesgo.
- UNE-EN 31010:2011. Gestión del riesgo. Técnicas de apreciación del riesgo.

10.3.12 Seguimiento

De cara a llevar a cabo el seguimiento de los riesgos, se deben realizar tareas periódicas. Primeramente, con una frecuencia semanal, el jefe de proyecto se reunirá con los responsables de cada riesgo para realizar una revisión de los mismos, así como una revaloración si tuviesen lugar cambios.

Asimismo, con una frecuencia mensual se realizará un análisis de nuevas amenazas que pudieran surgir en el contexto del proyecto, entre el jefe de proyecto y los *stakeholders*.

10.4 Artículo de investigación

El artículo ha sido enviado a la revista Semantic Web Journal (<http://semantic-web-journal.org/>) y, en el momento de entregar esta memoria, se encuentra en revisión. Una copia de tal artículo se incluye a continuación.

A dual approach to ShEx visualization with complexity management

Jorge Alvarez-Fidalgo ^{a,*} and Jose Emilio Labra-Gayo ^a

^a *Dept. of Computer Science, University of Oviedo, Spain*

E-mails: UO258524@uniovi.es, labra@uniovi.es

Abstract. Shape Expressions (ShEx) are used in various fields of knowledge to define RDF graph structures. ShEx visualizations enable all kinds of users to better comprehend the underlying schemas and perceive its properties. Nevertheless, the only antecedent (RDFShape) suffers from limited scalability which impairs comprehension in large cases. In this work, a visual notation for ShEx is defined which is built upon operationalized principles for cognitively efficient design. Furthermore, two approaches to said notation with complexity management mechanisms are implemented: a 2D diagram (Shumlex) and a 3D Graph (3DShEx). A comparative user evaluation between both approaches and RDFShape was performed. Results show that Shumlex users were significantly faster than 3DShEx users in large schemas. Even though no significant differences were observed for success rates and precision, only Shumlex achieved a perfect score in both. Moreover, while users' ratings were mostly positive for all tools, their feedback was mostly favourable towards Shumlex. By contrast, RDFShape and 3DShEx's scalability is widely criticised. Given those results, it is concluded that Shumlex may have potential as a cognitively efficient visualization of ShEx. In contrast, the more intricate interaction with a 3D environment appears to hinder 3DShEx users.

Keywords: Shape Expressions, Visual notation, UML, 3D, Cognitive load

1. Introduction

Shape Expressions (ShEx) [1] was proposed in 2014 as a language for RDF¹ data validation. By allowing to define RDF graph structures, it enables data producers and consumers to settle in a common ground and avoid inconsistencies. Since RDF brings together users from various branches of human knowledge, ShEx is employed in a variety of different contexts. E.g., ShEx is used to validate the RDF representation of FHIR², a standard for health care data exchange.

This implies that users do not necessarily have to be familiar with textual programming languages, resulting in a steep learning curve. One possible solution to such problem is the use of **visualizations**. They enable users to comprehend sheer amounts of data in an efficient manner and allows for better perception of emergent properties, errors and patterns [2].

The only precedent as far as ShEx is concerned is RDFShape [3], being capable of generating UML-like³ class diagrams for a subset of the language. Alas, it suffers from limited scalability as well as a degree of symbol overload which may affect its semantic transparency. Therefore, the information conveyed may be cognitively inefficient, particularly in larger use cases.

* Corresponding author. E-mail: UO258524@uniovi.es.

¹<http://www.w3.org/RDF/>

²<https://www.hl7.org/fhir/rdf.html>

³<https://www.omg.org/spec/UML/>

1 Other visualisations in the Semantic Web ecosystem formulate different solutions to the problem of developing a
2 comprehensible visual notation, with varying degrees of success. However, the aforementioned issue of scalability
3 -also referred to as complexity management- is rarely addressed. At most, a few -such as WebVOWL⁴- provide
4 automatic mechanisms for reducing the number of elements displayed, but no choice is given to the user about the
5 specifics.

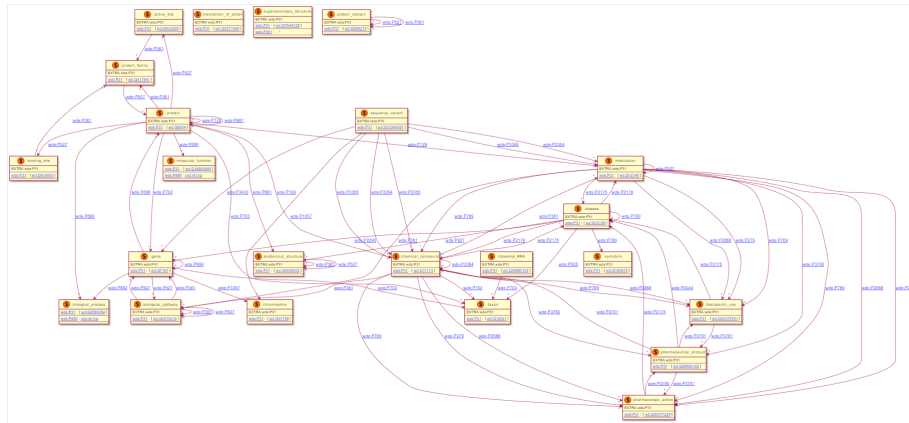
6 Thus, the main **contribution** of this work lies in the proposal of a visual notation for ShEx which aims to be
7 cognitively efficient -with an emphasis on complexity management-, analysing the perceptual implications of its
8 materialisation in both a 2D plane and 3D space.

9 The rest of the paper is structured as follows. A motivating example is provided in Section 2. In Section 3, back-
10 ground information about cognitive implications of visual design is provided, as well as the state of the art about
11 visualization tools in the Semantic Web. The proposed visual notation and both approaches to it are exposed in Sec-
12 tion 4. The implementation of the corresponding prototypes is discussed in Section 5. User evaluation methodology,
13 results and discussion are provided in Section 6. Finally, conclusions and future work are discussed in Section 7.

14 2. Motivating example

15 The Wikidata GeneWiki [4] project aims to use Wikidata as a semantic framework to manage and disseminate
16 biomedical data. To that end, it describes a knowledge graph about such entities and their relationships. Our ShEx
17 motivating example⁵ defines this graph structure. Given its abundant number of elements, it poses a challenge for
18 proper visualization.

19 RDFShape's visual representation (DOT)⁶ generates a class diagram with 23 classes and over 70 relationships (see
20 Fig. 1). Besides the cognitive implications of processing a large network -which will be discussed later-, common
21 scalability issues may be observed. Key sections of the diagram become filled with relationships and difficult to
22 discern between each other. Therefore, it is a suitable testing ground for testing complexity management mechanisms
23 and the cognitive efficiency of the visual notation.



24 Fig. 1. Genewiki ShEx visualization in RDFShape.

25 3. State of the Art

26 In this section, (i) cognitive implications of visual notation design and (ii) visualizations in the Semantic Web are
27 discussed.

28 ⁴<https://service.tib.eu/webvowl/>

29 ⁵<https://github.com/weso/sparkwdsb/blob/master/examples/genewiki.shex>

30 ⁶<https://rdfshape.weso.es/link/16520062624>

3.1. Cognitive implications of visual notation design

Cognitive load theory "is concerned with the manner in which cognitive resources are focused and used during learning and problem solving" [5]. It describes the impairment of understanding that takes place when learning procedures lead to further cognitive processes. A distinction is made between intrinsic and extraneous cognitive load; the former is due to the inherent complexity of the information, while the latter is generated because of the manner in which such information is presented. Those phenomena are related in such a way that the consequences of extraneous cognitive load may only be noticeable when dealing as well with intrinsic cognitive load caused by high element interactivity [6].

D. Moody describes in his **Physics of Notations** (PoN) theory a series of principles for designing cognitively effective visual notations [7]: *semiotic clarity*, *perceptual discriminability*, *semantic transparency*, *complexity management* (reduction of extraneous cognitive load), *cognitive integration*, *visual expressiveness*, *dual coding*, *graphic economy and cognitive fit*. In the last decade, PoN has become a widely used standard for notation design to the detriment of competing approaches [8].

Various criticisms have been stated about PoN. The operationalization of said principles ranges from objective measures -semiotic clarity is a 1:1 correspondence- to subjective evaluations -the "suggestion of meaning" implied by semantic transparency may only be determined by empirical means- [9]. This implies a degree of user involvement usually lacking in its application [10]. Subsequent proposals were made in order to improve such operationalization, either partially [11] or completely [12].

On a different note, the impact of 3D visualizations in cognitive load may be closely related to spatial ability [13]. I.e. subjects with high spatial ability perceive their cognitive load as low and vice versa. Further research shows that this effect is exacerbated when dealing with static visualizations; dynamic interactions providing a compensating effect for low spatial ability learners [14].

3.2. Visualizations in the Semantic Web

As stated in the introduction, **RDFShape** provides the only visualization currently available for Shape Expressions⁷. It generates a bidimensional graph in which UML-like boxes symbolize shapes and directional arrows represent references to other shapes. No interactive actions nor complexity management mechanisms are provided.

Further work has been carried out for the closely related Shapes Constraints Language (SHACL)⁸ in the form of visual editors. Arndt et al. implemented an Ontopad-based⁹ tool which allows for composing a SHACL visual data model [15]. Most of the interaction is done through a textual interface; new elements must be dragged into a canvas to become part of the visualization. Users may perform a few tasks on the visualization, such as linking properties.

Lieber et al. define both UML-based and VOWL-based visual notations to represent RDF constraints and implement them in **UnSHACLeD**, a SHACL visual editor [16]. Empirical tests showed no significant difference in error rates between the approaches. Nevertheless, the majority of users did prefer the VOWL-based notation. The authors acknowledge the need for complexity management mechanisms, but it is considered out of their scope.

VOWL [17] is a visual notation for representing OWL¹⁰ ontologies, with two implementations available: **WebVOWL** and **ProtégéVOWL**. **WebVOWL** makes use of a force graph, allowing for user interaction with the positioning of elements. Moreover, it provides a complexity management tool: a collapsing feature which reduces the number of elements on screen, even though it leaves no choice to the user over the specifics. Nonetheless, this does not prevent the overlapping of a large number of relationships between two nodes.

This overlapping issue has been a recurring problem in the history of visualizations in the semantic web. As far back as the year 2000, a possible solution emerged: to represent semantic graphs in three dimensions. With such purpose tools as **UNIVIT** [18] and **NV3D** were implemented; alas, their visual notations were too dependant on the arbitrary combination of visual variables (shape, color...) to efficiently convey complex data [19].

⁷<https://rdfshape.weso.es/shexInfo>

⁸<https://www.w3.org/TR/shacl/>

⁹<https://github.com/AKSW/OntoPad>

¹⁰<https://www.w3.org/TR/owl-features/>

A decade later, **X3D-UML** was proposed as a 3D UML implementation, particularly focused on state machine diagrams [20]. It consists of a number of interconnected planes in a tridimensional space, each one displaying a 2D UML diagram. Thus, it is rather an intermediate solution.

4. Proposal

This proposal consists of a UML class diagram-like visual notation in order to graphically represent ShEx, with two alternative approaches: a 2D diagram (**Shumlex**) and a 3D directed graph (**3DShEx**).

4.1. Visual notation

The proposed visual notation is displayed in Table 1. Its design rationale is structured according to PoN's principles, exposed hereunder.

4.1.1. Semiotic Clarity

Semiotic clarity is sacrificed in favour of both semantic transparency and graphic economy. Firstly, the notation incurs in a deliberate case of **symbol deficit**, since there exist a number of semantic constructs without a unique visual construct mapped to it. Node constraints are displayed textually inside shapes, much like attributes in UML classes. By doing so, it is expected to take advantage of the widely recognized UML class diagram notation to convey information to a broader audience.

Secondly, some semantic constructs employ the same visual construct (**symbol overload**) with textual differentiation. Since those semantic constructs are conceptually similar (e.g. conjunction and disjunction) it is hoped to achieve graphic economy without disrupting clarity.

4.1.2. Perceptual discriminability

In order to objectively ascertain the ease of discrimination between symbols, both a metric and a threshold of dissimilarity between two graphical symbols have to be defined [12]. To that end, the proposal from [11] is modified slightly in order to calculate the average of the following criteria: visual distance (VD), redundant coding (RC), perceptual pop-out (PPO) and textual differentiation (TD). These are normalized to an interval of [0, 1], in such a way that 0 denotes null discriminability and 1 compliance with all criteria. 0.5 is chosen as the threshold of dissimilarity.

For brevity's sake, the details of the modifications and calculations are exposed in Appendix B. The values obtained were $VD = 0.47$, $RC = 0.29$, $PPO = 1$ and $TD = 0.5$. This results in an average value of 0.57, therefore demonstrating the positive perceptual discriminability.

4.1.3. Semantic transparency

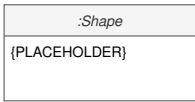
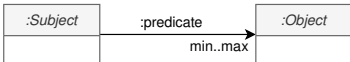
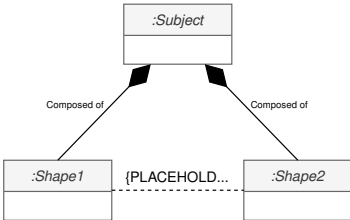
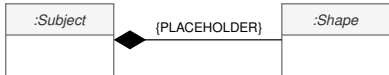
By shaping the notation to resemble a UML class diagram, the objective is to increase its semantic transparency, particularly to novice users. At least, semantic translucency is expected, so that the visual constructs provide a cue to its meaning by association. Nonetheless, this cannot be ascertained until a user evaluation is performed, given its aforementioned subjectivity.

4.1.4. Complexity management

Complexity management is approached in different ways. Shumlex takes inspiration from the modularization utility implemented by GraphQL Voyager¹¹, which by clicking on one of the classes highlights only that class and its relations with its neighbours, drastically reducing the others' visibility. A modification of this concept is proposed such that it is cumulative; that is, clicking on a second class does not change the focus to that class, but adds it to the highlighted set.

¹¹<https://apis.guru/graphql-voyager/>

Table 1
Proposed visual notation.

Feature	{PLACEHOLDER}	Visual representation	Example
TripleConstraint	<Property> <NodeConstraint> <Cardinality>;		:User { :name xsd:string ?; }
EachOf	<TripleConstraint>+		:User { schema:name xsd:string; schema:age xsd:string; }
Top-level Nodekind	nodeKind: <NodeKind>		:HomePage IRI
Extra	EXTRA <ValueSet>		:User EXTRA schema:name { schema:name xsd:string ; }
Closure	CLOSED		:User CLOSED { }
ShapeRef			:User { schema:worksFor @:Company ; }
ShapeAnd	AND		:User { schema:name xsd:string ; } AND { schema:age xsd:integer }
ShapeOr	OR		:User { schema:name xsd:string } OR { schema:givenName xsd:string ; }
OneOf	OneOf		:User { :name xsd:string ; (:givenName xsd:string +; :familyName xsd:string ;) ; }
ShapeNot	NOT		:NoName Not { schema:name . }
Labelled	Composed of		:User { \$.name (:name . ; :email IRI ;) }

As far as 3DShEx is concerned, an *intelligent zoom* [21] will be implemented. This implies additional interactivity besides the common zooming. By default, nodes will display only the shape identifier. When interacted with -opened-, nodes will expand to reveal the pertinent restrictions (*black boxing*). Further interaction -closing- will revert it to its initial state. As an additional complexity management tool, a simple collapsing function is proposed which on demand shows only the desired node and its neighbours. Both provide a layer of abstraction.

4.1.5. Cognitive integration

Given the fact that no multiple diagrams are used to represent a dataset, this principle does not apply.

4.1.6. Visual expresiveness

PoN builds upon Bertin's list of visual variables [22]: shape, texture, brightness, size, color, orientation and planar variables. The proposed notation makes use of the following: shape, texture and brightness. Hence, it lies in a middle ground between visual one-dimensionality and visual saturation. Moody claims that most diagrams in software engineering are visually one-dimensional [7] and therefore a higher degree of discriminability is achieved, as discussed earlier.

4.1.7. Dual coding

Every visual representation is complemented by text which provides a cue to its meaning.

4.1.8. Graphic economy

As previously stated, symbol deficit was introduced in order to reduce graphic complexity. Consequently, the number of graphical symbols is 4. Such is an inferior value to the 7 ± 2 processing capacity limit proposed by Miller [23], suggested in [12] as a reference for this principle.

4.1.9. Cognitive fit

The aim of the aforementioned binary approach is to maximize cognitive fit, each prioritizing different necessities. Shumlex aims to be closer to the UML class diagram spec, with the intention of being accessible to a wider audience not necessarily familiar with the technical details. Thus, all information is initially available as it would be in a common diagram, making for a more constraint-focused visualization in contrast to a relationship-focused 3DShex.

On the contrary, 3DShex is of a more experimental nature, which by interactively presenting the same information in a tridimensional space aims to further analyse the cognitive implications in its audience and the potential benefits it may bring. The details of shapes are concealed behind a layer of abstraction, thus giving greater importance to the diagram as a whole.

5. Implementation

In this section, the elaboration of the prototypes for both approaches to the visual notation is described.

5.1. Shumlex

In order to build the visualization Mermaid¹², a Javascript library for text-based generation of various diagrams, is used. Therefore, the architecture of the prototype is as follows:

1. A **conversion** engine which receives a ShEx input and generates the Markdown-like syntax that Mermaid requires. Given the fact that Mermaid does not accept a variety of symbols used in ShEx, it is necessary to use alternatives. For instance, the use of colons is not allowed; the prefixed term ":User" would have to be codified as "_User".
2. A **visualization** module which invokes the library with the previous outcome in order to generate a SVG. Once displayed, the sanitized texts are substituted by the original ones.
3. A **post-processing** module which implements the complexity management functionality. It assigns to every class in the diagram an event which, on click, lowers to a minimum the opacity of every element except that very class, its relationships and the targets of these. There exist a couple of exceptions: a) it won't obscure the already highlighted elements and b) it will reverse the effect if it has already been applied to said class. Furthermore, hovering any label will check the existence of the entity in Wikidata and display its meaning as a tooltip. The purpose of this is to increase comprehension of commonly used, Wikidata related Shape Expressions, in which shapes and predicates are semantically opaque (e.g. wd:Q42944 refers to CERN).

¹²<https://mermaid-js.github.io/mermaid/>

The application of Shumlex to the motivating example is shown in Fig. 2. Despite the fact that relationships are more spaced out than in RDFShape’s visualization, areas with high concentrations of elements remain cognitively overloaded. As shown in Fig. 3, the complexity management mechanism allows for a limited display of the desired components. In the provided example, focus is on `:medication`, thus highlighting its relationships with other shapes. This prototype is freely available at <http://www.weso.es/shumlex/>.

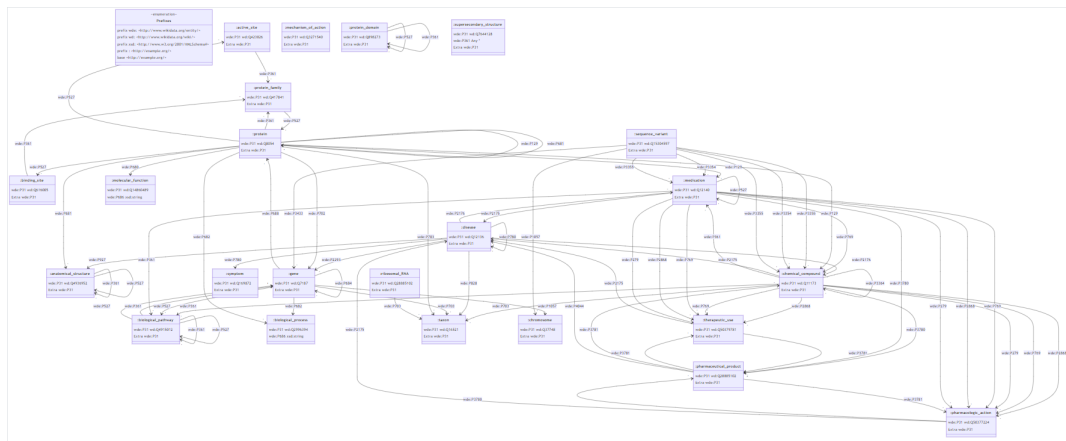


Fig. 2. Genewiki ShEx visualization in Shumlex.

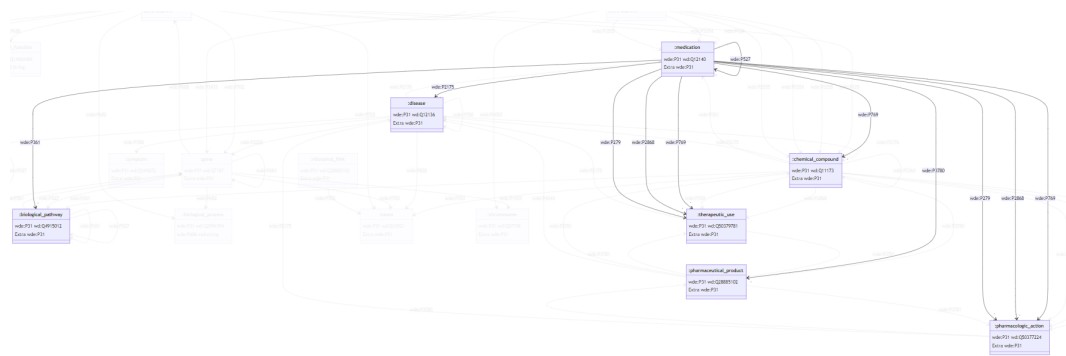


Fig. 3. Genewiki ShEx visualization with reduced complexity in Shumlex.

5.2. 3DShex

For the implementation of this prototype 3D Force Graph¹³ (3DFG), a NodeJS library to represent graph data in a tridimensional space, is used. The architecture of the prototype is as follows:

1. A **conversion** engine which receives a ShEx input and generates the JSON data that 3DFG requires. Besides the required parameters, additional information is included to facilitate the next phase.
 - (a) *List of constraints of a node.* Information to be displayed on demand, equivalent to the class attributes in UML.
 - (b) *Name and cardinality of a relationship.*

¹³<https://github.com/vasturiano/3d-force-graph>

- (c) *Curvature of a link.* Curved links allow for distinct relationships between a pair of nodes, while straight links only enable one to be displayed clearly. Therefore, references between shapes (:User :works @:Company) require the former since there may be any number of them. On the contrary, compositional relationships (ShapeAnd, ShapeOr, OneOf, ShapeNot and Labelled in Table 1) are unique to the source shape and thus are able to be represented by straight links.
- (d) *Arrow head.* As displayed in the notation, there are three possibilities: none, arrow or diamond.
- (e) *Rotation.* As previously mentioned, curved links allow for an arrangement free of the overlapping described in Section 3.2. However, by default, links are displayed in the same position. 3DFG allows for a rotation value -taking the node as the centre of a circumference- to be specified, but the calculations are left to the user. Hence, every link occurrence for each node pair is registered and the circumference is divided in equal parts. Moreover, the source of the link should be taken into account, since from the perspective of each circumference the angle will be different for a certain position (e.g. π in the source node equals to 2π in the target).
2. A **visualization** module which makes use of the previous information to invoke the library. HTML objects are utilized to build the nodes, so their contents can be customized as well as dynamic behaviour assigned (constraints are hidden by default). The following functionalities are enabled:
- (a) *Highlight on hover.* Both links and nodes possess this property; in the case of the latter, its neighbours and the corresponding relationships are emphasized as well. In links, moving particles are shown to reinforce the direction.
- (b) *Details.* When clicking a node, all its constraints are displayed in an expanded box. Another click reverts it to its original state.
- (c) *Collapsing.* Right clicking a node displays a reduced graph, composed of such node and its neighbours.
- (d) *Wikidata tooltips.* As in Shumlex.

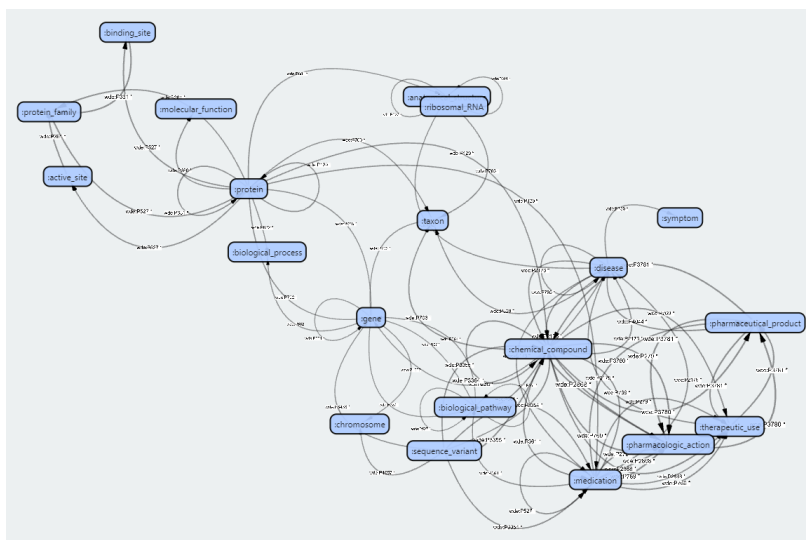


Fig. 4. Genewiki ShEx visualization in 3DShEx.

The application of 3DShEx to the motivating example is shown in Fig. 4. Even though static images do little for its comprehension -since it may be examined from any position- it is clear that clusters of highly interdependent shapes exceed working memory limits. The complexity management mechanism is thus applied to :medication yet again. As shown in Fig. 5, a much smaller graph is displayed, composed by the desired element and its neighbours.

This prototype is freely available at <http://www.weso.es/3dshex/>.

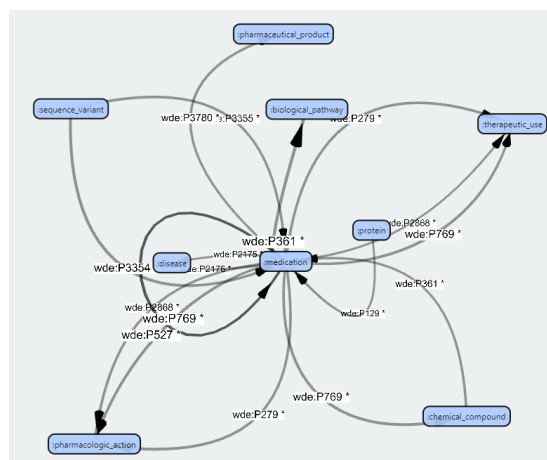


Fig. 5. Genewiki ShEx visualization with reduced complexity in 3DShEx.

6. Evaluation

In order to test the proposed notation, an experiment was carried out based in the one conducted in [24]. The methodology employed, the results obtained and their discussion are detailed in the following subsections. Datasets, questionnaires, manuals and anonymized results are freely available at <https://github.com/fidalgoLXXVI/shex-visualization-paper>.

6.1. Methodology

This user study follows a between-subjects design, in which each participant is exposed to a single tool and asked to perform a few measured tasks. Both a quantitative and a qualitative analysis are conducted. Hereunder, the methodology of this experiment is discussed in greater detail.

6.1.1. Procedure

The experiment is divided in the following steps:

1. **Preliminary questionnaire.** Subjects are inquired about background and self-assessment of relevant skills -such as knowledge in UML or spatial ability-.
2. **Tool description.** A brief manual is provided to participants which briefly describes the operation and features of the corresponding tool. The selected tools for the experiment were *RDFShape*, *Shumlex* and *3DShEx*. Thus, the different approaches to the notation may be compared to each other as well as to the existing solution.
3. **Main questionnaire.** A series of tasks on the test cases are requested to the participants. By means of the mandated tool, subjects must try to perform those while their interactions are measured by a timer. Each test case comprises the following tasks, which aim to ascertain the user's ability to navigate the diagram and comprehend the various semantic equivalences.
 - (a) Find a shape by name.
 - (b) Find a shape with a specific constraint.
 - (c) List within-node constraints of a shape.
 - (d) Find a reference between two shapes.
 - (e) Determine subject and object of a reference..
 - (f) List all neighbours of a shape.
4. **Follow-up questionnaire.** A number of questions based in the Likert scale are asked to the participants in order to perform the qualitative analysis. Those allow us to obtain a number of variables: *general satisfaction level*, *ease of use*, *learnability*, *semantic transparency*, *applicability*, *error proneness*, *scalability*, *complexity management*, *understanding of constraints* and *understanding of references*.

6.1.2. Sample

The sample consisted of 13 students of the MSc in Web Engineering at University of Oviedo. This experiment took place in the last day of a course in semantic web, where they were taught the basics of technologies such as RDF or ShEx. Most participants share a similar demographic as well as academic background, with a bachelor's degree in Computer Science. According to self-assessment results a) 92.3% have either medium or high knowledge of UML, b) 84.6% have basic knowledge of RDF, c) 69.2% have basic knowledge of ShEx and 15.4% declare no knowledge on the subject and d) 69% have high spatial ability while the rest declare medium spatial ability.

6.1.3. Test cases

Two test cases are used in this experiment. The first one is based on the WebIndex ShEx schema proposed in [25], "one of the earliest practical applications of ShEx". Modifications have been made in order to reflect all the features reflected in the visual notation, thus including logical operations and composition. The *OneOf* constraint is removed since RDFShape's current version hasn't implemented it yet. This schema features a few shape references, with greater focus on other node constraints.

The second one is the Genewiki schema, as featured in Section 2. It has approximately three times the shapes of the former and over 70 shape references, while other constraints are a scarce occurrence.

Hence, according to the cognitive implications layed out in Section 3.1, participants would be confronted with **distinct cognitive loads**. The first test case has a higher intrinsic load, given the greater inherent complexity of using complex semantic features such as conjunctions and composition while having few elements. On the contrary, the second test case has little implicit complexity -most are simple references to other shapes- but its large quantity of elements causes diagrammatic complexity upon display.

6.1.4. Threats to validity

Taking as reference the list of threats to both internal and external validity proposed in [26], the following have been identified:

Selection. Participants may share certain characteristics which predispose them towards the same results, especially given the common background. In order to address this, subjects are distributed randomly among the experimental groups so that those characteristics may be equally distributed.

Testing. Participants may become familiar with the test cases and remember responses for later tasks. In order to mitigate this, special care is taken to use different fragments of the schema and avoid repetitions.

Interaction of selection and treatment. Because of the limited variety of the participants, generalization to individuals of other contexts may not be possible. Hence, claims about the universality of the results must be restricted. However, given the highly specialized nature of the contribution, this issue is lessened.

6.1.5. Analysis

Both quantitative and qualitative results were collected and anonymised. From those, the following variables are calculated for each test case: **elapsed time**, **success rate** and **precision**. Elapsed time (T_c) is the total time spent for a given test case. Success rate (S_c) is calculated as the number of correct answers divided by the number of questions. Precision (P_c) is calculated as the division of minimum elapsed time of all participants by current student's elapsed time, multiplied by the success rate. This measure gives an insight on the swiftness of participants while taking into account their effectiveness. Hence, given a test case c and a student sn :

$$P_{c sn} := \frac{\min(\{T_{cs1}, \dots, T_{c sn}\})}{T_{c sn}} \cdot S_{c sn}$$

R 4.2.0 is used for the statistical analysis. Comparisons between the three groups are made by means of a One-Way ANOVA whenever assumptions are met, removing outliers if necessary. Otherwise, Kruskal-Wallis is used.

6.2. Results

Descriptive statistics of the quantitative results for the first test case are shown in Table 2. Shumlex mean scores are consistently better than RDFShape's, and those better than 3DShEx's. Nonetheless, those differences between the

three groups are not statistically significant for any of the variables: $F(2,8)=1.1$; $p=0.377$, $F(2,10)=1.67$; $p=0.236$ and $F(2,9)=1.29$; $p = 0.32$ for T, S and P respectively.

Table 2
Descriptive statistics for test case 1 results.

Measure	Group	\bar{x}	s	max	min
Elapsed seconds	3DShEx	256.2	66.55	355	210
	RDFShape	210.2	119.62	411	95
	Shumlex	196	95.63	302	73
Success rate	3DShEx	0.667	0.136	0.833	0.5
	RDFShape	0.7	0.139	0.833	0.5
	Shumlex	0.833	0.136	1	0.667
Precision	3DShEx	0.204	0.077	0.29	0.103
	RDFShape	0.311	0.197	0.64	0.118
	Shumlex	0.441	0.380	1	0.161

Descriptive statistics of the quantitative results for the second test case are shown in Table 3. Mean score comparisons show the same relationship between groups as before. However, in this case there are significant differences between the three groups in elapsed times ($H(2)=6.05$; $p=0.048$; $\eta^2=0.405$). Dunn post-hoc determined significant differences in elapsed times between Shumlex and 3DShEx ($p=0.014$). While 75% of Shumlex users achieve lower times than every RDFShape user, overall differences are not significant ($p=0.242$).

As success rate and precision are concerned, there are no significant differences between groups in the second test case ($H(2)=1.78$; $p = 0.41$ and $F(2,10)=2.43$; $p=0.137$).

Table 3
Descriptive statistics for test case 2 results.

Measure	Group	\bar{x}	s	max	min
Elapsed seconds	3DShEx	417.8	173.034	644	247
	RDFShape	265.6	106.746	456	204
	Shumlex	186.5	91.799	314	95
Success rate	3DShEx	0.583	0.096	0.667	0.5
	RDFShape	0.7	0.14	0.833	0.5
	Shumlex	0.708	0.21	1	0.5
Precision	3DShEx	0.154	0.0757	0.256	0.074
	RDFShape	0.268	0.074	0.357	0.174
	Shumlex	0.476	0.365	1	0.151

Descriptive statistics of the qualitative results are shown in Table 4 in Appendix A. Overall, user ratings are positive for all tools: 71.1% of answers express either high or very high level of approval. Tools score on average neutral or positive ratings for every measure, with the exception of Scalability which obtains neutral or negative ratings on average. Statistical analysis showed no significant differences between the three groups for any measure.

6.3. Discussion

Results for the first test case do not show any significant difference between groups for any of the metrics. This can be explained by the great variability in all groups -e.g. elapsed time for Shumlex ranges from $\sim 1m$ to $\sim 5m$ - as well as little difference between means.

Nonetheless, it should be noted that only one member of the Shumlex group achieved a perfect score both in success rate and precision. Overall, success rate is likely negatively influenced by the scarce theoretical knowledge of Shape Expressions which participants assessed. This would explain how a few simple tasks seem to cause general

1 confusion. Most notably, question 9 which involved pointing out the reference which connected two shapes got no
 2 correct answers from both 3DShEx and RDFShape groups, while most Shumlex users answered correctly. Given
 3 its uniqueness within the experiment, this particular difference in performance may be due to either an underlying
 4 cause or pure chance.

5 Thus, it may be only stated with certainty that **in cases with low diagrammatic complexity, there is no evidence**
 6 **of difference in performances between tools**. Since only the Shumlex group managed to complete all the proposed
 7 tasks and performed adequately in error-prone tasks, there may be need of further evaluation with larger samples to
 8 assess the potential influence of the tool in success rates and precision.

9 Regarding the second test case, **elapsed times show significant differences between groups** with a large effect
 10 size ($\eta^2=0.405$). Post-hoc results suggest that 3DShEx users require more time to perform tasks on large cases than
 11 Shumlex users. This may be explained by the combination of a novel navigable space with large diagrammatic
 12 complexity causing high cognitive load, thus exceeding working memory limits. This high cognitive load hypothesis
 13 is supported by the fact that the only 3DShEx user with lower spatial ability obtained the highest time in the
 14 experiment. As stated in Section 3.1, lower spatial ability may imply higher cognitive loads in 3D environments.
 15 Superior time performance for Shumlex users may be due to its closer resemblance to UML class diagrams, whose
 16 specification users claimed to be familiar with.

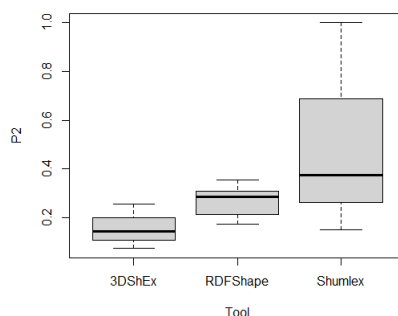


Fig. 6. Precision values by tool in the second test case.

31 Finally, results show no significant difference between groups at any of the variables in the subjective evaluation
 32 performed by the students. Average ratings are mostly positive or neutral if not. The sole exception in the qualita-
 33 tive analysis is the variable Scalability¹⁴, where the tools scored either neutrally or negatively. This suggests that
 34 users perceive the visualization tools to be of more use with small or medium schemas. Even though complexity
 35 management tools seem to be appreciated, extraneous cognitive load may be still excessive. This effect is probably
 36 exacerbated by their unfamiliarity with the language and the notation.

37 Oddly enough, while RDFShape scores a 3.8/5 in Complexity Management, when asked for feedback some of
 38 its users convey naught but dissatisfaction in this regard. *"As expected, the larger the use case, the more confusing*
 39 *the diagram"* or *"In very large graphs it is complicated to see the arrows that link entities in the central regions"*. In
 40 spite of those -expected- statements, neither rates it negatively.

41 In the light of such contradictions, there are several possible explanations. They might have not completely under-
 42 stood the statement to assess¹⁵, or they might have feared that too harsh an assessment would be detrimental to our
 43 interests (a RDFShape user even gives a perfect score to all variables but one). Mayhap it is merely a consequence
 44 of their lack of experience. Either way, it may seem like user feedback holds information of greater value to us.

45 Further analysis of user's comments unveils a similar perception of 3DShEx's complexity management. *"In small*
 46 *cases it is very useful, in large cases like the second one it is quite difficult to deal with"* and *"In the second case*
 47 *it was impossible to follow the relationships and to find the texts of the relationships for each shape"*. It cannot be
 48

14 "The tool is most useful in large use cases."

15 "The tool facilitates the understanding of complex areas."

concluded whether they actually used the complexity management mechanisms. However, if that is the case those may be unintuitive to the users; either way, the tool fails at meeting those needs.

By comparison, there is a single comment related to Shumlex: "*In the second case, [...] it can be a little complicated to discern the name of the relationship. You can select the shape from which it comes out to differentiate [the name] but it would be nice to be able to do it by clicking on it or hovering over it*". The contribution of the mechanism is appreciated while providing alternative solutions to that particular task.

As a summary of the qualitative analysis, **complexity management perception seems to be more favourable to Shumlex**, while the remaining variables appear to have a similar impact throughout the tools.

7. Conclusions and future work

A UML-based visual notation for ShEx has been proposed, which is built upon broadly used and operationalized principles. Moreover, said notation has been implemented in both 2D and 3D prototypes, named respectively *Shumlex* and *3DShEx*. Results of both qualitative and quantitative analysis lead to the following conclusions:

Efficiency of Shumlex. Even though Shumlex users mostly obtained better results than participants with other tools independently of the test case, the small sample size implies that those differences were not significant enough as to be able to generalize claims of efficiency. Nonetheless, the above together with receiving the most positive user feedback make us think that such universalization may be possible with further research. The absence of widespread complaints about complexity management -as it occurs in the others- is likely to be the result of the mechanisms put in place.

Cognitive overload in 3D environment. 3DShEx users were significantly slower than Shumlex users in a large use case. Furthermore, 3DShEx obtains the worst average ratings in success rate, precision and most qualitative variables. While differences are not statistically significant in those, it is considered likely that further research may provide a basis for confirmation. Lastly, despite having at their disposal a complexity management mechanism similar to that of Shumlex, user complaints are directed towards scalability. Given that intrinsic cognitive load is the same as Shumlex, it is concluded that interaction with the 3D environment is causing a greater extraneous cognitive load upon the user. The resulting cognitive overload frustrates the user to the detriment of their comprehension and proper use of the available features.

UML-like visual notation. By shaping the visual notation to resemble UML class diagrams, it was hoped to achieve a intuitive, transparent solution without forsaking efficiency. User evaluation of learning ease and semantic transparency is reasonably affirmative of such intent. On the other hand, its efficiency seems rather dependent on the manner in which the visual notation is presented.

Future work. It is considered that future efforts should be focused on Shumlex as the more promising approach. Analysis of user feedback suggests an extension of the complexity management capabilities as to support more specialized tasks. E.g., being able to select a single shape reference. Moreover, the inclusion of a search engine could be of assistance to users when navigating large schemas.

Appendix A. Qualitative analysis results

Table 4
Descriptive statistics for qualitative analysis results.

Measure	Group	\bar{x}	s	max	min
General satisfaction	3DShEx	3	0.82	4	2
	RDFShape	4.4	0.55	5	4
	Shumlex	4.25	0.96	5	3
Ease of use	3DShEx	3.75	1.5	5	2
	RDFShape	4.2	0.45	5	4
	Shumlex	4.5	1	5	3
Learnability	3DShEx	3.75	0.5	4	3
	RDFShape	4	1.22	5	2
	Shumlex	4.5	0.58	5	4
Semantic Transparency	3DShEx	3.75	0.5	4	3
	RDFShape	3.8	0.45	4	3
	Shumlex	4.25	0.5	5	4
Applicability	3DShEx	3.25	0.96	4	2
	RDFShape	4.2	0.45	5	4
	Shumlex	4	1.15	5	3
Error proneness	3DShEx	3.25	0.96	4	2
	RDFShape	3.6	1.14	5	2
	Shumlex	3.5	1.73	5	2
Complexity Management	3DShEx	3.5	1.29	5	2
	RDFShape	3.8	0.84	5	3
	Shumlex	3.25	0.96	4	2
Scalability	3DShEx	1.25	0.5	2	1
	RDFShape	3	1.22	5	2
	Shumlex	2.5	1.29	4	1
References	3DShEx	3.75	0.96	5	3
	RDFShape	4	0.71	5	3
	Shumlex	3.75	0.96	5	3
Constraints	3DShEx	3	0	3	3
	RDFShape	3.6	1.14	5	2
	Shumlex	3.75	0.96	5	3
Global	3DShEx	3.26	0.8	5	1
	RDFShape	3.86	0.82	5	2
	Shumlex	3.83	1.01	5	1

Appendix B. Metric of similarity

Given a notation N with G_N graphical symbols, the operationalization framework of Störrle et al. [11] proposes 4 criteria to ascertain perceptual discriminability: visual distance (VD), redundant coding (RC), perceptual pop-out (PPO) and textual differentiation (TD). These are normalized to an interval of $[0, 1]$, in such a way that 0 denotes null discriminability and 1 compliance with all criteria. Finally, the average discriminability for the notation is calculated.

B.1. Visual distance

Both the visual variable difference function $vvd(a,b)$ and the visual distance function $vd(g,h)$ are used as-is. Maximum value of $vd(g,h)$ is 1 and there are $|G_N^2|$ possible combinations of g,h . Hence, the average visual distance

function $VD(N)$ uses $|G_N^2|$ as a denominator so that 1 is the highest value possible. This does not take into account that whenever g equals h , $vd(g, h) = 0$, so the maximum value of the summation is $|G_N^2| - |G_N|$. Thus, the denominator is modified and the subtraction of the unit removed so that the values are normalized in the specified range.

$$VD(N) := \frac{1}{|G_N^2| - |G_N|} \sum_{g, h \in G_N} vd(g, h)$$

For instance, the visual distance between the graphical symbols *box* (b) and *directed arrow* (da) is as follows. As suggested, weights w are 7 for shape and 1 for the rest of visual variables. Those visual variables not used have $vvd = 0$, thus only shape, brightness and texture may be computed. In this particular case, shapes are in different main groups: lines and regions. Therefore, $vvd(v_{sh}(b), v_{sh}(da)) = 1$. Same for brightness, given that the colour of their main areas is completely opposite (black and white). However, they do have the same solid texture, hence $vvd(v_{tx}(b), v_{tx}(da)) = 0$.

$$vd(b, da) := \frac{1}{\|w\|} \sum_{i=1}^d w_i \cdot vvd(v_i(b), v_i(da)) := \frac{1}{14} (7 \cdot 1 + 1 \cdot 1 + 1 \cdot 0) := 0.57$$

This same procedure is repeated for all combinations of graphical symbols. The only new development is the comparison between shapes of the same basic group (i.e. arrows and lines) for which $vvd = 0.5$. Its results are shown in Table 5.

Table 5
Visual distance for all graphical symbol pairs.

	Box	Directed arrow	Diamond arrow	Dashed line
Box	0	0.57	0.57	0.64
Directed arrow	0.57	0	0.32	0.32
Diamond arrow	0.57	0.32	0	0.39
Dashed line	0.64	0.32	0.39	0

With all of the above, VD for our notation N may be finally calculated.

$$VD(N) := \frac{1}{16 - 4} (0.57 \cdot 4 + 0.32 \cdot 4 + 0.64 \cdot 2 + 0.39 \cdot 2) := 0.47$$

B.2. Redundant coding

Previous changes to the denominator apply to the $RC(N)$ function as well. The $vr(g, h)$ function is used as-is. Results are displayed in Table 6.

Table 6
Redundant coding for all graphical symbol pairs.

	Box	Directed arrow	Diamond arrow	Dashed line
Box	0	0.25	0.25	0.38
Directed arrow	0.25	0	0.25	0.25
Diamond arrow	0.25	0.25	0	0.38
Dashed line	0.38	0.25	0.38	0

$$RC(N) := \frac{1}{16 - 4} (0.25 \cdot 8 + 0.38 \cdot 4) := 0.29$$

B.3. Perceptual popout

Each graphical symbol has at least a unique value in one visual variable. Taking into account the *shape* variable alone fulfils this criterion. Once more, the subtraction is removed so that the best possible value is 1. Therefore, the function is as follows:

$$PPO(N) := \frac{1}{4}(1 \cdot 4) := 1$$

B.4. Textual differentiation

As stated in Section 4.1.1, PoN does not consider visual constructs that make use of textual differentiation to convey distinct meanings as different graphical symbols (symbol overload). Therefore, a modification is made to this criterion so that it does not measure the proportion of graphical symbols which only differ by textual cues. The proportion of graphical symbols which convey several concepts by textual differentiation is evaluated instead.

$$TD(N) := 1 - \frac{2}{4} := 0.5$$

References

- [1] E. Prud'hommeaux, J.E. Labra Gayo and H. Solbrig, Shape expressions: an RDF validation and transformation language, *Proceedings of the 10th International Conference on Semantic Systems - SEM '14* (2014). doi:10.1145/2660517.2660523.
- [2] C. Ware, *I. Foundations for an Applied Science of Data Visualization*, in: *Information visualization: Perception for design*, Morgan Kaufmann, an imprint of Elsevier, 2021.
- [3] J.E. Labra Gayo, D. Fernández Álvarez and H. García González, RDFShape: An RDF Playground Based on Shapes, *ISWC 2018 Posters & Demonstrations* (2018).
- [4] S. Burgstaller-Muehlbacher, A. Waagmeester, E. Mitiraka, J. Turner, T. Putman, J. Leong, C. Naik, P. Pavlidis, L. Schriml, B.M. Good and et al., Wikidata as a semantic framework for the Gene Wiki Initiative, *Database* **2016** (2016). doi:10.1093/database/baw015.
- [5] P. Chandler and J. Sweller, Cognitive load theory and the format of instruction, *Cognition and Instruction* **8**(4) (1991), 293–332–. doi:10.1207/s1532690xci08042.
- [6] J. Sweller, Cognitive load theory, learning difficulty, and instructional design, *Learning and Instruction* **4**(4) (1994), 295–312–. doi:10.1016/0959-4752(94)90003-5.
- [7] D. Moody, The “physics” of notations: Toward a scientific basis for constructing visual notations in software engineering, *IEEE Transactions on Software Engineering* **35**(6) (2009), 756–779–. doi:10.1109/tse.2009.67.
- [8] D. van der Linden and I. Hadar, A systematic literature review of applications of the physics of notations, *IEEE Transactions on Software Engineering* **45**(8) (2019), 736–759–. doi:10.1109/tse.2018.2802910.
- [9] D. van der Linden, A. Zamansky and I. Hadar, How cognitively effective is a visual notation? on the inherent difficulty of operationalizing the physics of notations, *Enterprise, Business-Process and Information Systems Modeling* (2016), 448–462–. doi:10.1007/978-3-319-39429-9_28.
- [10] D. van der Linden and I. Hadar, User involvement in applications of the pon, *Lecture Notes in Business Information Processing* (2016), 109–115–. doi:10.1007/978-3-319-39564-7_1.
- [11] H. Störrle and A. Fish, Towards an operationalization of the “physics of notations” for the analysis of visual languages, *Lecture Notes in Computer Science* (2013), 104–120–. doi:10.1007/978-3-642-41533-3_7.
- [12] D. van Der Linden, A. Zamansky and I. Hadar, A framework for improving the verifiability of visual notation design grounded in the physics of notations, *2017 IEEE 25th International Requirements Engineering Conference (RE)* (2017). doi:10.1109/re.2017.37.
- [13] T. Huk, Who benefits from learning with 3D models? the case of Spatial ability, *Journal of Computer Assisted Learning* **22**(6) (2006), 392–404–. doi:10.1111/j.1365-2729.2006.00180.x.
- [14] T.N. Höffler, Spatial ability: Its influence on learning with visualizations—a meta-analytic review, *Educational Psychology Review* **22**(3) (2010), 245–269–. doi:10.1007/s10648-010-9126-7.
- [15] N. Arndt, A. Valdestilhas, G. Publio, A. Cimmino, K. Höffner and T. Riechert, A Visual SHACL Shapes Editor Based On OntoPad, in: *Joint Proceedings of the Semantics co-located events: Poster&Demo track and Workshop on Ontology-Driven Conceptual Modelling of Digital Twins co-located with Semantics 2021*, International Conference on Semantic Systems Proceedings, Amsterdam, Netherlands, 2021. <http://ceur-ws.org/Vol-2941/paper16.pdf>.
- [16] S. Lieber, B. Denbsp;Meester, P. Heyvaert, F. Brückmann, R. Wambacq, E. Mannens, R. Verborgh and A. Dimou, Visual notations for viewing RDF constraints with unshacled, *Semantic Web* (2021), 1–36–. doi:10.3233/sw-210450.

- 1 [17] S. Lohmann, S. Negru, F. Haag and T. Ertl, Visualizing ontologies with Vowl, *Semantic Web* 7(4) (2016), 399–419–. doi:10.3233/sw-
2 150200.
- 3 [18] B. Le Grand and M. Soto, Information management - Topic Maps visualization (2000).
- 4 [19] V. Geroimenko, B. Le Grand and M. Soto, *Topic Maps, RDF Graphs, and Ontologies Visualization*, in: *Visualizing the semantic web: XML-based internet and information visualization*, Springer, 2004, pp. 66–69–.
- 5 [20] P. McIntosh, X3D-UML: User-Centred Design, Implementation and Evaluation of 3D UML Using X3D, *RMIT University* (2009).
- 6 [21] L. Bartram, R. Ovans, J. Dill, M. Dyck, A. Ho and W.S. Havens, Contextual Assistance in User Interfaces to Complex, Time-Critical
7 Systems: The Intelligent Zoom (1994).
- 8 [22] J. Bertin, *Graphics and graphic information-processing*, W. de Gruyter, 1981.
- 9 [23] G.A. Miller, The magical number seven, plus or minus two: Some limits on our capacity for processing information., *Psychological Review*
10 63(2) (1956), 81–97–. doi:10.1037/h0043158.
- 11 [24] H. García-González, I. Boneva, S. Staworko, J.E. Labra-Gayo and J.M. Cueva Lovelle, ShExML: Improving the usability of heterogeneous
12 data mapping languages for first-time users, *PeerJ Computer Science* 6 (2020). doi:10.7717/peerj-cs.318.
- 13 [25] J.E. Labra Gayo, E. Prudhommeaux, I. Boneva and D. Kontokostas, *Validating RDF data*, Morgan amp; Claypool Publishers, 2018.
- 14 [26] J.W. Creswell and J.D. Creswell, *Research design: Qualitative, quantitative, and mixed methods approaches*, SAGE Publications, Inc,
15 2018.
- 16
- 17
- 18
- 19
- 20
- 21
- 22
- 23
- 24
- 25
- 26
- 27
- 28
- 29
- 30
- 31
- 32
- 33
- 34
- 35
- 36
- 37
- 38
- 39
- 40
- 41
- 42
- 43
- 44
- 45
- 46
- 47
- 48
- 49
- 50
- 51