

Tipo de artículo: Artículo original
Temática: Ingeniería de Proyectos
Recibido: 25/05/2018 | Aceptado: 10/09/2018

Descubrimiento de resúmenes lingüísticos para ayuda a la toma de decisiones en gestión de proyectos

Discovering linguistic summaries for help in project management decisions

Iliana Pérez Pupo¹ [0000-0003-1433-0601]*, Osvaldo Santos Acosta¹ [0000-0002-6315-4782], Roberto García Vacacela² [0000-0002-1834-6806], Pedro Y. Piñero Pérez¹ [0000-0002-7635-8290], Esther C. Ramírez¹ [0000-0003-1587-9524]

¹ Grupo de Investigaciones en Gestión de Proyectos, Universidad de las Ciencias Informáticas, La Habana, Cuba. {[iperez](mailto:iperez@uci.cu), [osantos](mailto:osantos@uci.cu), [ppp](mailto:ppp@uci.cu), [estherc](mailto:estherc@uci.cu)}@uci.cu

² Universidad Católica De Santiago de Guayaquil, Ecuador. roberto.garcia@cu.ucsg.edu.ec

* Autor para correspondencia: iperez@uci.cu

Resumen

Las técnicas de sumarización lingüística de datos agrupan un conjunto de algoritmos de minería de datos útiles para descubrir relaciones intrínsecas presentes en los datos. Estas relaciones son presentadas en lenguaje natural para facilitar la toma de decisiones en la temática objeto de estudio. Existen diferentes técnicas para la generación de resúmenes entre los que se destacan el uso de consultas de bases de datos aplicable en los resúmenes más sencillos y el uso de meta-heurísticas en los resúmenes de mayor complejidad. En este trabajo se propone un algoritmo para la generación de resúmenes lingüísticos a partir de datos heterogéneos y tomando como base la generación de reglas de asociación. Además, se emplean las medidas propuestas por Zadeh para la evaluación de los resúmenes combinados con técnicas de aprendizaje activo. Finalmente, se aplica la técnica propuesta para la toma de decisiones en gestión de proyectos y se discute acerca de las decisiones tomadas a partir de los resúmenes obtenidos.

Palabras clave: gestión de proyectos, reglas de asociación, resumen lingüístico de datos

Abstract

The linguistic data summarization consists on data mining techniques used to discover intrinsic relationships present in the data. These techniques generate linguistic summaries from discovered relationships. There are different algorithms to generate summaries, the simplest summaries can be generated by using standard query languages.

Other authors built summaries by using metaheuristics such as genetic algorithms. This paper presents a new linguistic data summarization techniques based on combination of algorithms to generate association rules, fuzzy logic and active learning. Summaries are evaluated by a combination of T values proposed by Zadeh and active learning techniques. Finally, the proposed technique is applied in project management context. The paper discusses different decisions taken form linguistic summaries.

Keywords: association rules, degree of truth, linguistic data summarization, project management

Introducción

La interpretación correcta de los datos es una limitación aún latente en muchas empresas y organizaciones que afecta al proceso de toma de decisiones y selección de estrategias. Un escenario particular de esta situación lo constituyen las organizaciones orientadas a proyectos. En este tipo de organizaciones, a pesar de los esfuerzos por mejorar la eficiencia y la eficacia en la gestión, persisten numerosas dificultades en los procesos de toma de decisiones. Los estudios continuos realizados por el *Standish Group International* (International, 2014) (Clancy, 2014), en 50.000 proyectos de todo el mundo, desde pequeños hasta gigantescos proyectos de reingeniería muestran una evolución de esta problemática, ver Figura 1. Se observa que los proyectos renegociados han repuntado en los últimos cinco años mientras que los satisfactorios han declinado, incluso, por debajo de los cancelados.

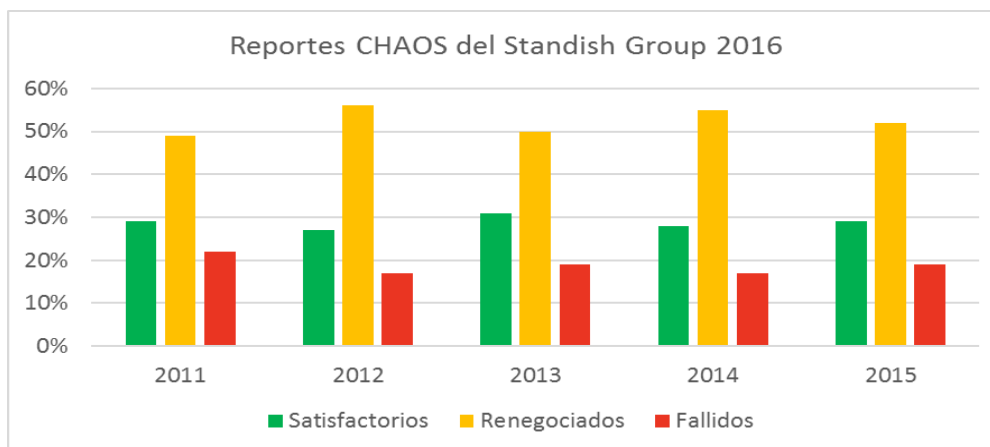


Figura 1: Evolución de los proyectos según los reportes del Standish Group entre el 2011-2015 (Clancy, 2014)

Como se puede ver, entre las causas fundamentales del fracaso en los proyectos se señalan: la mala gerencia, las insuficiencias en los procesos de planificación y de control y seguimiento (Pacelli, 2004)(Delgado, Miguel; Ruiz, M.

Dolores; Sánchez, Daniel; Vila, 2014)(Villavicencio, 2016). Estas causas pueden ser mitigadas si se cuenta con técnicas para el descubrimiento de conocimiento y el análisis de resúmenes históricos de los datos en forma lingüística, de modo que estos resúmenes puedan ser comprensibles para los decisores y a su vez puedan ser utilizados para facilitar tareas como el análisis de decisiones, predicción o pronóstico (Gomes and Romão, 2016)(Boran et al., 2016) (Díaz et al., 2011)(Piñero et al., 2014).

En este contexto, surge la sumarización lingüística de datos (SLD, también conocido como LDS del inglés: *linguistic data summarization*) como una de las técnicas de descubrimiento de conocimiento descriptivo con un enfoque interesante y prometedor para producir resúmenes a partir de datos numéricos utilizando lenguaje natural. Con ella las organizaciones podrían resolver el dilema de “datos ricos, información pobre” presente en la toma de decisiones. Para un mejor análisis de resultados de investigaciones en este campo y su aplicación en las organizaciones orientadas a proyectos, se realizó una búsqueda bibliográfica en las bases de datos SCOPUS, Google Scholar, Thomson Routers, IEEE, Semantic Scholar, Library Genesis y Microsoft Academic correspondiente al tema “*linguistic data summarization*” (ver Tabla 1). En esta búsqueda se detectó un grupo de insuficiencias en las técnicas existentes que dejan un campo abierto a la investigación, las cuales se muestran a continuación.

Tabla 1: Resultados de la búsqueda bibliográfica.

"Linguistic Data Summarization"	IEEE	Thomson Routers	Google Scholar	Semantic Scholar	SCOPUS	Library Genesis	Microsoft Academic
Últimos 5 años	62	45	22	7	5	2	1

- En la búsqueda realizada no se encuentran aplicaciones de las técnicas de sumarización lingüística de datos en las organizaciones orientadas a proyectos.
- Existe necesidad de construcción de algoritmos que permitan aplicar técnicas de aprendizaje con refuerzo y auto-aprendizaje en los procesos de generación de los resúmenes.
- Existen insuficiencias en la eficiencia y la eficacia de los métodos existentes para la obtención de los resúmenes, en particular en los métodos basados en meta-heurísticas.

Los resúmenes lingüísticos son clasificados en dos grandes grupos (Zadeh, 1983) (Donis-Díaz et al., 2014):

$Qy's\ are\ S \quad T(La\ mayoría\ de\ los\ empleados\ tienen\ bajo\ salario) = 0.7$

$QRy's\ are\ S \quad T(La\ mayoría\ de\ los\ empleados\ jóvenes\ tienen\ bajo\ salario) = 0.7$

Kacprzyk y Zadrozny han definido las siguientes clasificaciones de la sumarización lingüística (llamadas protoformas) teniendo en cuenta los elementos que ya se conocen y los que se buscan.

Tabla 2: Clasificación de la sumarización lingüística (Zadeh, 2002) (Díaz et al., 2011) (Kacprzyk and Zadrozny, 2005).

Tipo	Protoforma	Conocido	Duda	Comentarios
0	$QRy's\ are\ S$	Todo	T	
1	$Qy's\ are\ S$	S	Q	Resúmenes simples a través de consultas ad-hoc
2	$QRy's\ are\ S$	S R	Q	Resúmenes condicionales a través de consultas ad-hoc
3	$Qy's\ are\ S$	$Q\ S^{Estructura}$	S^{Valor}	Resúmenes sencillos orientados a valores
4	$QRy's\ are\ S$	$Q\ S^{Estructura},\ R$	S^{Valor}	Resúmenes orientados a valores condicionales
5	$QRy's\ are\ S$	Nada	Q R S	Reglas difusas generales

Para medir la calidad del resumen se han creado varios indicadores (Zadeh, 1983) los cuales se relacionan a continuación:

- Grado de verdad (T1): mide el grado de verdad del resumen considerando el grado de pertenencia de los objetos al sumariador del resumen y su correspondencia con el cuantificador del resumen en cuestión.
- Grado de imprecisión (T2): indica cuan impreciso o vago es el resumen considerando el conjunto de valores posibles para el sumariador y cómo estos valores son cubiertos por los conjuntos borrosos que conforman la variable del sumariador.
- Grado de cobertura (T3): indica la frecuencia relativa de objetos de la base de datos, que cumplen simultáneamente con los filtros establecidos y con el valor del sumariador del resumen. En esencia, es una medida que muestra el grado de cubrimiento de la base de datos por el resumen lingüístico.
- Grado de adecuación (T4): mide en qué grado el resumen lingüístico es útil y se calcula en función de combinar la frecuencia respecto a la ocurrencia del sumariador en los objetos de la base de datos y la cobertura. Esta medida indica que un resumen cuyo sumariador esté presente en el 100% de los objetos de la base de datos tendrá un alto valor de verdad, pero un muy bajo valor del T4, el cual será 0.

- Longitud de un resumen (T5): este indicador se refiere a la longitud del resumen, es decir, la cantidad de variables involucradas en él, bajo el principio de que un resumen muy largo puede ser de difícil comprensión por los interesados.
- Valor integrado de los criterios de calidad (T6): este indicador integra al resto de los valores de T. Se pueden usar diferentes operadores para esta integración, como por ejemplo la media, la media ponderada o un operador OWA (Kacprzyk and Zadrożny, 2018).

Más adelante, estos valores de T se usarán para medir el grado de calidad de los resúmenes obtenidos en esta propuesta.

Metodología computacional

En esta sección se propone un algoritmo para la construcción de resúmenes lingüísticos de datos, generando los mismos a partir de reglas de asociación. Como parte de las actividades que deben ser realizadas para preparar los datos para la construcción de resúmenes se encuentran:

1. Preparación del conjunto de datos (D).
2. Definición de la variable lingüística del cuantificador (Q) formada por los siguientes conjuntos borrosos: “Muy pocos” [0,0,0.17]; “Pocos” [0,0.17,0.33]; “Algunos” [0.17,0.33,0.5]; Alrededor del 50% [0.33,0.5,0.67]; Muchos [0.5,0.67,0.83]; “La mayoría” [0.67,0.83,1.0] y “casi todos” [0.83,0.83,1.0].
3. Definición de los términos lingüísticos de las variables que describen el dominio de los datos.
4. Establecimiento de las parejas de funciones T-norma / Co-norma a emplear en el proceso de evaluación de los resúmenes con los diferentes valores de T.

A continuación, se presenta el algoritmo propuesto.

Algoritmo AprioriUnificatorLDS

Entradas:

D: conjunto de datos para el análisis.

FuzzyVar: relación de variables lingüísticas con sus valores lingüísticos y funciones de pertenencias.

RulesParameters: parámetros para generar las reglas de asociación (soporte y confianza).

Q: variable lingüística de los cuantificadores de los resúmenes.

Object: objeto 'y' de la base de datos.

Par_{T-S_norma}: operadores de agregación, par T-norma y S-norma.

Umbral: umbral (ϵ) utilizado para el cálculo de las T y para cuantificar.

p_value. Valor que se utiliza en algunos operadores de agregación.

Paso 1. Inicio

Paso 2. $D_f = \text{Fuzzify}(D, \text{FuzzyVar})$ // Transformación de los datos D en borrosos

a. memberships = CalculateMemberships // Genera matriz del grado de pertenencia de cada valor a su etiqueta lingüística

Paso 3. Rules = Extract_rules(Df, RulesParameters) // Genera reglas de asociación usando a priori

a. Eliminar reglas donde:

(rule.antecedent == rule.consequent) or rule.consequent.empty? or rule.consequent.nil?

Paso 4. UnifiedRules = UnifierRules(Rules) // Se unifican las reglas extraídas en el paso anterior, por cada regla:

*a. rule_i.antecedent == rule_{i+1}.consequent and
rule_i.consequent == rule_{i+1}.antecedent*

b. rule_i.consequent == rule_{i+1}.consequent

Paso 5. Summaries = CreateSummaries(UnifiedRules, Object, Q) // Genera los resúmenes lingüísticos, donde por cada regla:

a. summarizer = rule.consequent

b. filter_values = rule.antecedent

c. quantifier = GenerateQuantifier(memberships, Q)

d. Summary = GenerateSummary(sumarizer, filter_values, quantifier, Object)

e. Summaries_T = TCalculator(Summary, Par_{T-S_normas}, p_value) // Se calculan los valores T de cada resumen

Paso 6. LS_{List} = Reorder(Summaries_T) // Reordenar los resúmenes lingüísticos en función del valor del T6.

Paso 7. LS_{List} = TranslateSummary (LS_{List}) // Traducir resúmenes empleando diccionario de variables.

Paso 8. LS_{List} = AprendizajeActivo (LS_{List})

// Refinar los resúmenes aplicando Aprendizaje Activo.

Paso 9. Devolver LS_{List} // lista ordenada de los resúmenes lingüísticos y mejorada por los expertos.

Paso 10. Fin

En el *Paso 1* del algoritmo, el usuario debe proveer dos ficheros en formato ‘csv’: uno con las variables y sus valores para la carga de los datos numéricos que se deseen procesar, y el otro con la configuración de los algoritmos que se emplean. En el *Paso 2* el algoritmo transforma el conjunto de datos (D) en datos borrosos (D_f), utilizando las variables lingüísticas con sus funciones de pertenencias correspondientes ($FuzzyVar$). En este paso también se genera una matriz (*memberships*) que contiene, para cada valor de las variables, el grado de pertenencia de ese valor a sus etiquetas lingüísticas.

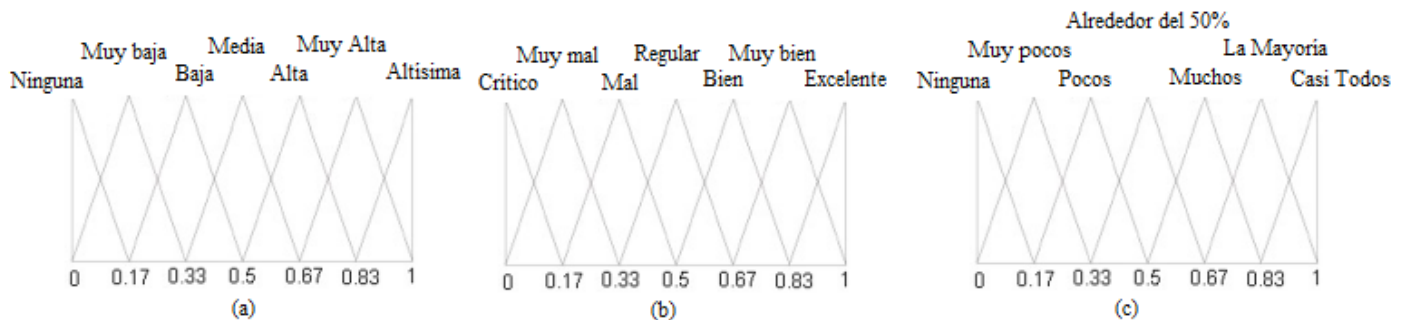


Figura 2: Ejemplos de variables lingüísticas empleadas para describir las variables, FuzzyVar. La variable lingüística (c) constituye la variable de los cuantificadores.

El *Paso 3* genera las reglas de asociación (utilizando apriori) a partir del conjunto de datos borrosos (D_f). Al generar las reglas se eliminan aquellas que:

- El antecedente sea el mismo consecuente (`rule.antecedent == rule.consecuent`)
- Reglas que no presenten consecuentes (`rule.consecuent.empty?` or `rule.consecuent.nil?`)

En el *Paso 4* se unifican las reglas generadas. Para ello, por cada regla se aplican los siguientes operadores de implicación:

- El bicondicional $p \leftrightarrow q$, que leemos "p sí y solo sí q" o "p es equivalente a q".
- Cuando los antecedentes de una regla es subconjunto de otra con el mismo consecuente.

En el *Paso 5*, del conjunto de reglas que fueron unificadas, se genera un resumen por cada una de ellas, donde por cada regla:

- Su consecuente es el sumador.
- Su antecedente es el filtro (R).

- Se calcula el cuantificador en función del grado de pertenencia del sumador y el filtro aplicando un operador T-norma, dividido por el grado de pertenencia del filtro.
- Con estos elementos se genera el resumen.
- Finalmente se calculan los valores T propuestos por Zadeh y explicados en la sección anterior. En particular el valor del T6 se calculó usando el operador “media ponderada” con los pesos (0.25, 0.16, 0.24, 0.25, 0.1) para los valores de (T1, T2, T3, T4, T5) respetivamente.

En el *Paso 6* a partir de los valores del T6 de cada resumen, se ordenan de mayor a menor, quedando en una posición superior los resúmenes de mejor calidad respecto al T6

En el *Paso 7*, una vez construidos los resúmenes, estos en ocasiones son difíciles de interpretar considerando el nombre literal de las variables en la base de datos. En este sentido, para garantizar la mejor legibilidad de los resúmenes, se incluyó un módulo de traducción que permite asociar a cada variable una descripción. El módulo traductor recibe como parámetro un fichero en forma de diccionario y sustituye literalmente el nombre de la variable por su significado, mejorando la legibilidad del resumen. Se muestra a continuación un ejemplo de resumen y su traducción.

Resumen original obtenido:

Muchos “proyectos” con (“trtr” Perfecto) o (“cant_rrhh_eval_m” Alta) o (“cant_comp_alta” Alrededor 50%) o (“cant_comp_baja” Alrededor del 50%) o (“icd ” Bien) o (“tptr” Mal) o (“tptp” Perfecto) o (“cant_rrhh_eval_b” Media) o (“cant_rrhh_eval_r” Media) tienen “ie” Mal.

Resumen traducido por el módulo traductor:

Muchos “proyectos” con (“tiempo real del trabajo realizado” Perfecto) o (“cantidad de recursos humanos evaluados de mal” Alta) o (“cantidad de recursos humanos con competencia alta” Alrededor del 50%) o (“cantidad de recursos humanos con competencia baja” Alrededor del 50%) o (“índice de calidad del dato” Bien) o (“tiempo planificado del trabajo realizado” Mal) o (“tiempo planificado del trabajo planificado” Perfecto) o (“cantidad de recursos humanos evaluados de bien” Media) o (“cantidad de recursos humanos evaluados de regular” Media) tienen “Índice de ejecución” Mal.

En el *Paso 7* se evalúan los resúmenes por un colectivo de expertos teniendo en cuenta los valores de T y la relevancia del resumen en correspondencia con el problema concreto, usando técnicas de computación con palabras.

En el *Paso 9*, en función del resultado de la evaluación de los expertos, se obtiene la lista final de los resúmenes.

Resultados y discusión

Para la aplicación y validación del algoritmo, se tomaron datos registrados sobre gestión de proyectos (Hechavarría et al., 2017) (Rivero and Pérez Pupo, 2018). Los datos relacionan indicadores de gestión con evaluación de proyectos, donde intervienen 25 variables y 8430 instancias. Las variables son:

Tabla 3: Variables involucradas en la selección de los resúmenes

Variables	Temática
cant_comp_alta, cant_comp_media, cant_comp_baja,	Cantidad de personas por niveles de competencias
cant_rrhh_eval_b, cant_rrhh_eval_m, cant_rrhh_eval_r	Cantidad de personas evaluadas en un nivel de desempeño determinado
fondo_tiempo, tiempo_planificado, tiempo_trabajado, tptp, tptr, trtr	Variables asociadas al fondo de tiempo de los recursos humanos y al tiempo planificado y real en cada corte de proyectos.
Icd, iref (calidad del dato y calidad) ie, ire, irp (control de la gestión de tiempo) irl (adquisiciones) irha, irhe, irhf, irht, irrh (recursos humanos)	Indicadores evaluados en cada corte de proyecto. Permiten evaluar al proyecto respecto a la eficiencia y la eficacia, cubriendo las áreas de conocimiento de la gestión de proyectos.
eval_fuzzysystem_advanced_01	Indicador de la evaluación final del proyecto

Una vez ejecutado el algoritmo se generaron 154 resúmenes lingüísticos, de ellos, fueron analizados en el contexto de este trabajo aquellos que involucraron los siguientes grupos de variables: evaluación de proyectos: (*eval_fuzzysystem_advanced_01*), evaluación de los recursos humanos (*cant_rrhh_eval_b* y *cant_rrhh_eval_m*) y competencia de los recursos humanos (*cant_comp_alta* y *cant_comp_baja*)

De este análisis se redujo a 88 resúmenes, los cuales fueron sometidos a un proceso de evaluación de la relevancia a partir de criterios de expertos empleando técnicas de computación con palabras modelo escala de ordinales. Los expertos evaluaron cada resumen considerando los valores de las seis *T* calculadas y la relevancia para la toma de decisiones, empleando el siguiente conjunto de términos: $LBTL = \{Muy\ bajo, Bajo, Medio, Alto, Muy\ alto\}$. Se recopilan las preferencias de los 7 expertos. Como resultado de esta evaluación 88 resúmenes fueron evaluados de “*Muy Baja*” relevancia, 20 resúmenes fueron evaluados de “*Baja*” relevancia, 14 resúmenes fueron evaluados de “*Media*” relevancia, 9 fueron evaluados de “*Alta*” y 6 fueron evaluados de “*Muy Alta*”. Se presentan a continuación los resúmenes con mayor relevancia identificados por los expertos y un análisis de las decisiones que son tomadas a partir de los mismos.

Resumen 1. *Muchos* “proyectos” con ("cantidad de recursos humanos con competencias altas" *Alrededor del 50%*) o ("cantidad de recursos humanos con competencias bajas” *Alrededor del 50%*) o ("cantidad de recursos humanos evaluados de mal" *Alta*) o ("cantidad de recursos humanos evaluados de bien" *Media*) o ("cantidad de recursos humanos evaluados de regular" *Media*) tienen “Índice de ejecución” *Mal*

Calidad del resumen: $T(0.79, 0.54, 0.69, 0.13, 0.031, 0.49)$

En este caso se debe señalar que el desempeño de los recursos humanos tiene un impacto significativo en la ejecución de los proyectos. Y que numerosos proyectos evaluados de mal comparten el mal desempeño de los recursos humanos. La organización objeto de estudio en este caso, tomó la decisión de concentrarse en los equipos de trabajo en cuyos proyectos se concentraban las principales dificultades.

Resumen 2. *Alrededor del 50%* de los “proyectos” con "cantidad de RH evaluados de mal" *Alta* tienen "tiempo real de trabajo real" *Perfecto*.

Calidad del resumen: $T(0.95, 0.47, 0.47, 0.002, 0.5, 0.476)$

Este resumen tiene un elevado valor de verdad (0.95) e indica que *Alrededor del 50%* de los proyectos tienen una “*Alta*” cantidad de recursos evaluados de “*Mal*” y sin embargo, tienen el tiempo real del trabajo real “*Perfecto*”. Esto significa que aparentemente, en estos proyectos hay una declaración falsa de la cantidad de trabajo real realizado. La decisión a tomar por la gerencia de la organización respecto a estos proyectos es verificar que se declaren correctamente los tiempos reales dedicados y que los jefes de equipo y gerentes de proyectos velen por la calidad de la información registrada en el sistema de información.

Resumen 3. *Alrededor del 50%* de los “proyectos” con ("aprovechamiento de los RH" *Regular*) o ("cantidad de RH evaluados de bien" *Baja*) o ("desempeño de los RH" *Regular*) o ("fondo de tiempo" *Media*) tienen "eficacia" *Mal*.

Calidad del resumen: $T(0.84, 0.77, 0.44, 0.08, 0.063, 0.47)$

Este resumen indica que existe una marcada influencia de los recursos humanos en la calidad del proyecto. Cuando los recursos humanos están mal y están dedicando poco tiempo al desarrollo de las actividades uno de los criterios que se afecta inmediatamente es la calidad. En este caso la decisión a tomar por parte de la gerencia de la organización es preocuparse por la calidad y el tiempo que los recursos humanos dedican al desarrollo.

Resumen 4. *Alrededor del 50%* de los “proyectos” con (“índice de ejecución” *Mal*) o (“cantidad de RH evaluados de bien” *Baja*) o (“evaluación de proyecto” *Regular*) o (“cantidad de RH evaluados de regular” *Baja*) tienen “ejecución en proceso del proyecto” *Mal*.

Calidad del resumen: $T(0.50, 0.86, 0.55, 0.18, 0.062, 0.45)$

En este resumen, en el 50% de los proyectos que están mal, presentan dificultades con su producción en proceso. Esta situación acarreará mayores conflictos con los clientes y se deben tomar medidas de control emergentes para evitar estas situaciones.

Resumen 5. *Alrededor del 50%* de los “proyectos” con “cantidad de RH evaluados de mal” *Media* tienen “eficacia” *Mal*.

Calidad del resumen: $T(0.78, 0.77, 0.47, 0.13, 0.5, 0.51)$

Este resumen indica que en los proyectos donde está mal el desempeño de los recursos humanos, está mal la calidad; además, la gerencia de la organización debe tomar las medidas asociadas con la elevación de la calidad a partir de aplicar convenientemente recompensas y penalizaciones al personal. Otra medida puede ser la contratación de nuevo personal más capacitado o capacitar al personal actual en aras de elevar la calidad.

Resumen 6. En *pocos* “proyectos” con “tiempo trabajado” *Perfecto* tienen “competencia baja de RH” *Alrededor del 50%*.

Calidad del resumen: $T(0.51, 0.26, 0.23, 0.18, 0.5, 0.32)$

Este resumen indica la dependencia de las competencias laborales con la eficiencia en el trabajo, se reporta que existen pocos proyectos con una excelente gestión de tiempo que tienen el 50% del personal con bajas competencias. Y como decisión, se debe continuar elevando las competencias del personal en las organizaciones como factor de éxito en la calidad y la eficiencia en el trabajo.

Otros resúmenes obtenidos permitieron identificar que el indicador de índice de ejecución de las tareas influye significativamente en la evaluación del proyecto. Además, permitieron detectar proyectos donde hay problemas en la evaluación de recursos humanos y otros en los que hay una mala distribución de los recursos humanos a partir de sus competencias. Respecto al valor de los T , se debe señalar que el resultado de la agregación de estos valores no arrojó el mismo orden de relevancia de los resúmenes que el orden declarado por los expertos en proyectos. Los autores de este artículo consideran que la identificación de otras medidas para evaluar los resúmenes puede constituir en trabajos futuros de investigación.

Conclusiones

- Con la aplicación de las técnicas de sumarización lingüística de datos en el escenario de aplicación, se logró elevar la efectividad en la gestión de proyectos y se identificaron situaciones que ayudaron a la toma de decisiones.
- La combinación de algoritmos para la construcción de reglas de asociación con las técnicas de sumarización lingüística de datos, facilita la obtención de resúmenes, pero tiene como particularidad que genera un número elevado de resúmenes que requieren ser filtrados en etapas posteriores de la búsqueda.
- Los autores de este artículo consideran que se debe continuar investigando en medidas para la evaluación de los resúmenes lingüísticos, en particular se deben construir valores de T que tengan en cuenta particularidades de los diferentes dominios de aplicación y que acerquen la evaluación de los resúmenes a la valoración de los expertos en cada dominio.

Referencias

- Boran, F.E., Akay, D., Yager, R.R., 2016. An overview of methods for linguistic summarization with fuzzy sets. *Expert Syst. Appl.* 61, 356–377. doi:10.1016/j.eswa.2016.05.044
- Clancy, T., 2014. The standish group chaos report. *Proj. Smart.*
- Delgado, Miguel; Ruiz, M. Dolores; Sánchez, Daniel; Vila, M.A., 2014. Fuzzy quantification: a state of the art. *Fuzzy Sets Syst.* 242. doi:10.1016/j.fss.2013.10.012
- Díaz, C.A.D., Pérez, R.B., Morales, E.V., 2011. Using Linguistic Data Summarization in the study of creep data for the design of new steels, in: *Intelligent Systems Design and Applications (ISDA), 2011 11th International Conference on.* IEEE, pp. 160–165.
- Donis-Díaz, C., Muro, A., Bello-Pérez, R., Morales, E.V., 2014. A hybrid model of genetic algorithm with local search to discover linguistic data summaries from creep data. *Expert Syst. Appl.* 41, 2035–2042. doi:http://dx.doi.org/10.1016/j.eswa.2013.09.002
- Gomes, J., Romão, M., 2016. Improving project success: A case study using benefits and project management. *Procedia Comput. Sci.* 100, 489–497.
- Hechavarría, C.C.R., Pupo, I.P., Pérez, P.Y.P., Huergo, R.H.B., 2017. Proceso de limpieza de datos en la construcción del repositorio para investigaciones en gestión de proyectos. *Process of cleaning data in the construction of the repository for research in project management.*
- International, T.S.G., 2014. *The CHAOS Manifesto.* The Standish Group International, Incorporated.

- Kacprzyk, J., Zadrozny, S., 2005. Linguistic database summaries and their protoforms: towards natural language based knowledge discovery tools. *Inf. Sci.* 173, 281–304. doi:10.1016/j.ins.2005.03.002
- Kacprzyk, J., Zadrozny, S., 2018. Reaching Consensus in a Group of Agents: Supporting a Moderator Run Process via Linguistic Summaries, in: *Soft Computing Applications for Group Decision-Making and Consensus Modeling*. Springer, pp. 465–485.
- Pacelli, L., 2004. *The Project Management Advisor: 18 major project screw-ups, and how to cut them off at the pass*. Pearson Education.
- Piñero, P., Pérez Pupo, I., Menéndez, J., 2014. Sistema de Información para la Gestión de Organizaciones Orientadas a Proyectos. Presented at the V Congreso Iberoamericano de Ingeniería de Proyectos, Loja Ecuador.
- Rivero, C.C., Pérez Pupo, I., 2018. Repositorio de Investigaciones en Gestión de Proyectos, Bases de Datos de Evaluación de Proyectos. Maestría en Gestión de Proyectos / Seminario de Investigaciones, Departamento de Investigaciones en Gestión de Proyecto, Universidad de las Ciencias Informáticas, La Habana, Cuba.
- Villavicencio, N.E., 2016. Modelo integrado para la mejora de la productividad en organizaciones orientadas a proyectos de tecnologías de la información. Tesis para optar al grado de: Máster en Diseño, Gestión y Dirección de Proyectos. Fundación Universitaria Iberoamericana.
- Zadeh, L.A., 1983. A computational approach to fuzzy quantifiers in natural languages. *Comput. Math. Appl.* 9, 149–184.
- Zadeh, L.A., 2002. A prototype-centered approach to adding deduction capability to search engines-the concept of protoform, in: *Intelligent Systems, 2002. Proceedings. 2002 First International IEEE Symposium*. IEEE, doi: 10.1109/IS.2002.1044219, pp. 2–3.