

Aplicación de técnicas de text mining para analizar las interacciones de los estudiantes en el proceso de aprendizaje de gestión de proyectos

Rubén Olarte-Valentín¹, Ana González-Marcos¹, Fernando Alba-Elías¹, Joaquín Ordieres-Meré²

Universidad de La Rioja, Logroño, Spain
ruben.olarte@alum.unirioja.es,
{ana.gonzalez, fernando.alba}@unirioja.es
Universidad Politécnica de Madrid, Madrid, Spain
j.ordieres@upm.es

Abstract. En este artículo se presenta la aplicación de técnicas de *text mining* para analizar la comunicación online de estudiantes que trabajan juntos en un mismo proyecto, con el fin de identificar la aparición de problemas en el desarrollo de la experiencia de aprendizaje en gestión de proyectos. Los datos empleados en este estudio son los mensajes que los estudiantes intercambiaron a través de las herramientas de comunicación existentes en la plataforma web empleada específicamente para el desarrollo de los proyectos. Los resultados obtenidos muestran una baja variabilidad en la terminología empleada por los estudiantes, lo que, sumado a un problema recurrente en minería de datos, esto es, la existencia de un conjunto de entrenamiento desbalanceado, convierte el proceso de clasificar mensajes en un problema complejo. Al mismo tiempo, se ha identificado como crucial la incorporación de términos propios del dominio de estudio en el análisis de sentimiento basado en diccionarios.

Keywords: gestión de proyectos · comunicación asincrónica · text mining

1 Introducción

El espacio europeo de educación superior supone un proceso de transformación sustancial de los aspectos organizativos, pedagógicos y metodológicos de la transmisión del conocimiento (Consejo de la Unión Europea, 2004). En este contexto de cambio, el nuevo modelo educativo se centra en el aprendizaje y desarrollo de competencias, surgiendo, de este modo, la necesidad de reajustar la metodología de enseñanza a las exigencias requeridas.

En línea con esta nueva concepción de la educación superior, existe un creciente interés en el proceso de enseñanza y aprendizaje en gestión de proyectos (Berggren y Söderlund, 2008; Ojiakoa et al., 2011). En este sentido, se han propuesto diferentes iniciativas en el ámbito educativo de gestión de proyectos basadas en la metodología de “aprender haciendo” (Alba-Elías et al., 2014; González-Marcos et al., 2016; Rooij,

2009; Thomas y Mengel, 2008; Turner et al., 2000). De hecho, dado que el aprendizaje de gestión de proyectos es, en sí mismo, un aprendizaje experiencial, es necesario proveer a los estudiantes de un ambiente en el que puedan actuar como directores de proyecto en la ejecución de un proyecto.

Trabajar en un proyecto significa trabajar en equipo, y un equipo de proyecto puede ser visto como un grupo social donde los miembros del equipo interactúan sí, comparten intereses y tienen el objetivo común de finalizar el proyecto con éxito. Por ello, en este trabajo, basado en el marco de aprendizaje presentado en (Alba-Eliás et al., 2013), donde los estudiantes asumen tres roles diferentes – ejecutivo (EX), director de proyecto (PM) y miembro del equipo (TM) –, todos ellos definidos en PRINCE2™ (Projects IN a Controlled Environment) (Office Of Government Commerce, 2009), el principal objetivo es analizar las interacciones entre los estudiantes a través de sus conversaciones online, con el fin de identificar la aparición de problemas en el desarrollo de la experiencia de aprendizaje. Más concretamente, este trabajo analiza la capacidad de las técnicas de *text mining* para extraer información acerca del carácter positivo o negativo de los mensajes intercambiados por los miembros de cada equipo de proyecto. El fin último de este análisis es contar con herramientas que permitan mejorar la identificación temprana de problemas en el desarrollo de la experiencia, para tratar de prevenirlos y poder actuar, incluso, antes de que estos se produzcan.

Diferentes estudios han explorado el comportamiento social de los alumnos durante la comunicación mediada por ordenador (CMC) (Choa et al., 2007; George y Leroux, 2002). De hecho, algunos estudios realizados a gran escala han identificado algunas diferencias significativas entre la comunicación asincrónica y la comunicación síncrona, las cuales parecen ser sutiles y se encuentran principalmente al realizar análisis cualitativo en grupos más pequeños (Hrastinski, 2008):

- La comunicación asincrónica es preferible cuando el propósito es discutir ideas complejas.
- Los alumnos disfrutaban más las discusiones síncronas por su carácter más social, pero varios estudios han identificado participaciones más concisas y menos profundas.

Este trabajo se centra en las conversaciones asíncronas, ya que tienden a estar mejor estructuradas y desarrolladas que la comunicación síncrona (Girasoli y Hannafin, 2008), además de ofrecer a los miembros del equipo de proyecto el tiempo necesario para examinar y reflexionar sobre un tema antes de que formalicen su contribución (Hammond, 2005) o proporcionar información relacionada con una parte del trabajo realizado.

2 Metodología

2.1 Participantes y procedimiento

Los participantes de este estudio fueron un total de 214 alumnos de grado y máster, matriculados en asignaturas de gestión de proyectos, en dos universidades distintas y dos cursos académicos diferentes. Los mensajes analizados se obtuvieron de una de las

herramientas de comunicación proporcionadas por el software de gestión de proyectos empleado en el desarrollo de la experiencia (<http://www.project.net>): los foros de discusión.

A través de estos foros, a los que tienen acceso todos los miembros del proyecto, estos pueden consolidar pensamientos e ideas, a la vez que comparten comentarios con otros miembros del proyecto. En esta aplicación particular, los foros de discusión también se utilizaron para informar a los miembros del proyecto responsables de un entregable o documento específico de que el trabajo solicitado ha sido finalizado. De este modo, la persona responsable tiene la posibilidad de responder para proporcionar información acerca del trabajo realizado, tanto para aceptarlo como para solicitar cambios imprescindibles antes de su aprobación.

2.2 Pre-procesado de los datos y análisis exploratorio

Limpieza del texto.

Antes de proceder a la utilización de los datos de texto, es conveniente aplicarles una serie de transformaciones básicas para permitir que su manipulación posterior sea más simple. El primer paso consiste en convertir todas las letras del texto a minúsculas. Además, como labores previas de adecuación del texto se sustituyen las vocales acentuadas por sus equivalentes sin acentuar, se suprimen los signos de puntuación, los caracteres numéricos, saltos de línea y de párrafo, las tabulaciones y cualquier espacio intermedio sobrante.

Extracción y selección de características.

Para poder analizar y utilizar el texto en un proceso de clasificación, es necesario proveerlo de una estructura que lo represente de manera cuantitativa. Una forma para conseguirlo es convertirlo en una matriz documento-término, en la que cada fila representa un texto y cada columna equivale a un término único de éste. La matriz recoge la cantidad de veces que cada término aparece en un determinado texto.

Antes de obtener la matriz documento-término, es posible aplicar varios tipos de funciones de extracción de características, con las que reducir los términos o variables que compondrán la matriz documento-término.

La primera operación consiste en eliminar del conjunto de términos aquellos de carácter común y escaso significado semántico. Éstos reciben el nombre de *stop words*, existiendo listas de estos términos en varios idiomas. En concreto, la lista empleada incluye determinantes, conjunciones, preposiciones y el conjunto de conjugaciones verbales de los verbos ser, estar, haber y tener.

La segunda operación de reducción de términos se conoce como *stemming* y su objetivo es reducir los términos a su raíz léxica. Así, aunque se pierde parte del significado concreto de la palabra original, aquellos términos cuyos lexemas coinciden tras su transformación pasan a agruparse como una sola variable.

Una vez reducido el número de variables, la matriz documento-término es sometida a un proceso de ponderación denominado TF-IDF (*Term Frequency-Inverse Document Frequency*) (Russell, 2013). El método se divide en dos fases y expresa el producto

entre los términos *Term Frequency* (TF) e *Inverse Document Frequency* (IDF), donde el primero equivale a la importancia que un término posee en un documento concreto y el segundo da un valor de la relevancia del término en el conjunto de textos. El TF (ecuación (1)) es considerado como el número de veces que un término aparece en un determinado texto, pero para conseguir un valor relativo a la longitud del texto es aconsejable tener en cuenta su cantidad total de palabras.

$$TF = \frac{\text{n}^\circ \text{ veces por texto que aparece el término}}{\text{n}^\circ \text{ de términos del texto}} \quad (1)$$

El IDF (ecuación (2)) es una medida para amplificar el valor de aquellos términos poco frecuentes en el conjunto de documentos y, a su vez, reducir el valor de aquellos que aparecen de manera frecuente. El logaritmo se añade para reducir el resultado a un rango limitado, ya que su función es la de ser un factor de escala del TF.

$$IDF = 1 + \log \left(\frac{\text{n}^\circ \text{ de textos del corpus}}{\text{n}^\circ \text{ de textos en los que aparece el término}} \right) \quad (2)$$

2.3 Clasificación

Para poder emplear algoritmos de clasificación supervisada, es necesario que los textos estén etiquetados como pertenecientes a una determinada categoría, que en este caso se ha elegido como carácter positivo o negativo, siendo esta valoración una base para permitir detectar de manera automática aquellos mensajes que impliquen algún tipo de problema surgido en el transcurso de la experiencia de manera automática y generar alertas sobre la posible necesidad de intervención.

La clasificación de los mensajes se ha realizado utilizando como criterio para detectar los mensajes negativos aquellos que cumplen las siguientes especificaciones:

- Problemas con las tolerancias de tiempo: problemas de urgencia, fuera de plazo, etc.
- Mensajes de aprobación de documentación que incluyen que ésta se ha debido a la falta de tiempo (a la vez que se considera que la calidad no era la apropiada).
- Quejas de cada figura del proyecto en dirección a sus superiores.
- Menciones al plan de excepción.
- Asunción de responsabilidades que ocasionan aprobaciones forzadas para poder contabilizar el trabajo realizado por los miembros subordinados.

Por su parte, los mensajes positivos son aquellos que muestran un desarrollo previsible dentro de la dinámica de la asignatura.

De entre las técnicas de clasificación supervisada empleadas de forma habitual en los análisis de *text mining*, como, por ejemplo, máquinas de vectores soporte (SVM), análisis de sentimiento basado en diccionarios (*lexicon-based sentiment analysis*) o el clasificador Naive Bayes, en este trabajo se han empleado las dos primeras.

Máquinas de vectores soporte (SVM).

El fundamento del algoritmo SVM es la capacidad de separar en dos clases un conjunto de datos (Vapnik, 1998). Se pueden aplicar a problemas de clasificación y de

regresión en una infinidad de dominios. Debido a su versatilidad cuentan con diversas opciones de configuración.

Análisis de sentimiento basados en diccionarios (*lexicon-based sentiment analysis*).

La clasificación de sentimiento basada en un léxico es la técnica más simple empleada para medir la polaridad de un sentimiento dentro de un texto, pero a la vez es un método robusto y de considerable dificultad debido a la gran variedad de sentidos que una palabra puede tener dentro de un texto (Danneman, 2014). Esta técnica requiere de un diccionario de términos o lexicón en el que cada palabra tiene asociada una puntuación de polaridad, positiva o negativa. La suma de las puntuaciones de todas las palabras de un texto que estén incluidas en el lexicón da como resultado el carácter positivo o negativo del mismo.

Hay dos tipos de diccionarios predefinidos: los genéricos y los de dominio específico. Además, existen dos formas de ampliar los diccionarios: la primera requiere de un diccionario para añadir nuevas palabras relacionadas con las *corpus* (texto analizado), como sinónimos y antónimos, y la segunda se basa en la adición de términos clave extraídos del propio *corpus* debido a su ausencia dentro del léxico original. Estos dos enfoques sirven para generar diccionarios contruidos empíricamente con el objetivo de calibrar el tipo de sentimiento subyacente en un texto.

3 Resultados y discusión

3.1 Pre-procesado de los datos y Análisis Exploratorio

Descripción de los datos.

Los datos utilizados en este estudio son un conjunto de 6668 mensajes intercambiados por los estudiantes durante el desarrollo de los proyectos en los que participan. Los textos están divididos en dos grupos. El primero, que consta de 5388 mensajes o textos formados por una cantidad de 8377 términos diferentes, se corresponde con los datos utilizados en el entrenamiento de los modelos de clasificación. Estos datos provienen en su conjunto de un único curso académico. El segundo está formado por 1280 mensajes con un número de términos de 3940 y procede íntegramente de otro curso académico diferente.

Extracción y selección de características.

Tras la limpieza inicial de los textos (conversión a minúsculas, supresión de tildes, signos de puntuación, números, saltos de línea y espacios sobrantes), se efectuó la eliminación de los *stop words*, lo que supuso reducir el número de términos en 139.

Para observar la influencia de la eliminación de este tipo de términos de la colección de textos, en la Fig. 1 se muestran dos *wordclouds* en los que se representan los términos más habituales pertenecientes a los textos originales (izquierda) y los textos procesados (derecha), con un tamaño proporcional a su frecuencia. De esta manera, es posible apreciar cómo la supresión de términos conforme al criterio de su bajo interés semántico influye notablemente en la distribución de frecuencias de los términos del conjunto de

textos del estudio, de forma que si en el grupo de mensajes en bruto los términos más comunes eran "que", "del" y "los", tras el primer paso de limpieza los más frecuentes son ahora "documento", "aprobación" y "saludo", todos con un carácter más representativo para el grueso de mensajes.

El siguiente paso consiste en aplicar el proceso de *stemming*. El resultado de su aplicación es una reducción de los términos hasta 4117, lo que supone una gran ventaja puesto que una dimensionalidad demasiado alta implica procesos clasificación y de modelización de datos más lentos y menos precisos.

Una vez procesados los textos, es necesario estructurarlos para generar a partir de ellos datos cuantificables. En primer lugar, se crea la matriz documento-término, aplicando sobre ella la ponderación TF-IDF, último paso de la fase de pre-procesado realizado en este trabajo. El resultado es bastante interesante, ya que hasta el momento, los diagramas de frecuencias mostraban una clara polarización de los términos hacia ciertas palabras muy recurrentes fruto de la especificidad del contexto en que son generados los mensajes. Tras la ponderación, las diferencias entre el uso de unos términos y otros se reduce considerablemente, alterando incluso el orden de las variables de mayor frecuencia de uso (Fig. 2).

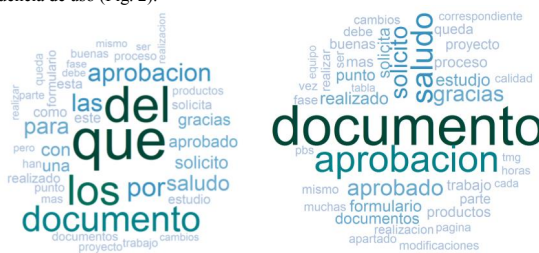


Fig. 1. Wordclouds de los textos originales (izquierda) y de los textos tras eliminar los *stopwords* (derecha).



Fig. 2. Wordclouds tras el *stemming*, sin (izquierda) y con (derecha) ponderación.

Una primera conclusión del análisis realizado es que la naturaleza de los datos hace que existan una serie de términos muy recurrentes, lo que muestra una baja variabilidad en el conjunto de textos intercambiado entre los estudiantes.

3.2 Clasificación

Como se ha indicado en el apartado de metodología, es necesario proceder a etiquetar los mensajes para poder utilizarlos como base de una modelización supervisada.

El resultado de dicha categorización en dos grupos es un gran número de mensajes positivos (5210) frente a una reducida cantidad de negativos (178), lo que plantea un problema de balanceo de datos a la hora de su utilización como datos de entrenamiento. Apparentemente, no existen grandes diferencias en cuanto a los términos de mayor frecuencia. Para conocer con mayor detalle sus características, se recurre de nuevo a la ponderación TF-IDF, que muestra cómo las palabras más relevantes en uno y en otro caso sí se pueden considerar bastante diferenciadas (Fig. 3 y Fig. 4).

Diagrama de Frecuencias de los Términos

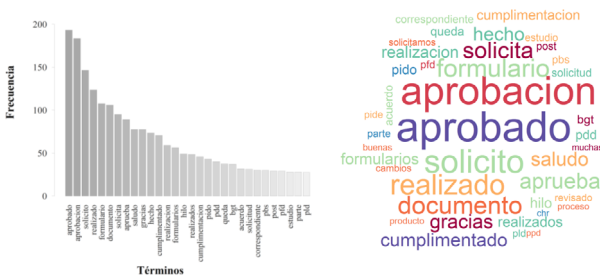


Fig. 3. Diagramas de caracterización de los textos positivos tras la ponderación TF-IDF.

Máquinas de vectores soporte (SVM).

Dado el elevado número de mensajes de carácter positivo en relación con la cantidad de mensajes negativos existente, la aplicación directa del modelo de clasificación SVM generó un modelo incapaz de predecir algún tipo de mensaje etiquetado como negativo en el conjunto de test. Para lograr mejorar el proceso de aprendizaje, se emplearon las siguientes estrategias (Table 1):

- Reducir el número de variables mediante el establecimiento de un límite en el número de veces en que aparece una palabra en el conjunto de los textos. En este caso, aunque la identificación de los mensajes negativos en el conjunto de test fue muy mala, si se observó un cambio tendencia.

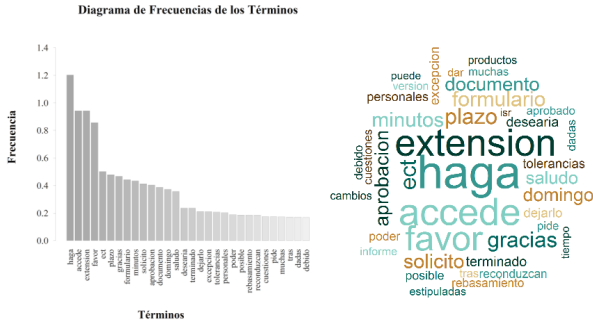


Fig. 4. Diagramas de caracterización de los textos negativos tras la ponderación TF-IDF.

- Balancear el conjunto de datos mediante una selección aleatoria de los textos pertenecientes a la clase “positivo”. En este caso, se logró mejorar ligeramente la clasificación de los mensajes negativos, pero a costa de reducir la cobertura para los mensajes de carácter positivo.

Table 1. Resultados del conjunto de test con los clasificadores SVM entrenados con distintos conjuntos de entrenamiento.

Conjunto de entrenamiento	Método de evaluación	Clase “Positivo”	Clase “Negativo”	Total
Conjunto inicial, sin términos poco frecuentes	Precisión	0.941	0.029	0.889
	Cobertura	0.944	0.027	0.892
	F-Measure	0.942	0.028	0.890
Conjunto balanceado, sin términos poco frecuentes	Precisión	0.954	0.102	0.905
	Cobertura	0.810	0.356	0.784
	F-Measure	0.876	0.159	0.835

Análisis de sentimiento basado en diccionarios (lexicon-based sentiment analysis).

La clasificación mediante el uso de un diccionario es considerado un método supervisado sin necesidad de utilización de datos etiquetados. Sin embargo, para una utilización más precisa de este sistema, se combinó la categorización de datos generada con anterioridad con los resultados de la clasificación basada en un léxico en castellano (Pérez-Rosas et al., 2012). De esta manera, la adecuación de los términos que forman el diccionario se efectúa mediante la inclusión de términos procedentes del propio conjunto de textos. En concreto, del grupo de entrenamiento. En la tabla 2 se observa que la aplicación de este procedimiento, denominado “diccionario original”, ofrece algunas mejoras con respecto a los modelos SVM anteriores. En concreto, se consigue mejorar

la precisión de la clase “negativo” sin reducir la cobertura de la clase “positivo”, aunque, como contrapartida, la cobertura de la clase “negativo” se ve reducida.

Tras el análisis de estos resultados y en vista de la descompensación existente entre la clasificación de resultados positivos y negativos, se procedió a la incorporación al diccionario de términos extraídos de los mensajes etiquetados como negativos para intentar aumentar la sensibilidad del diccionario frente a la aparición en los mensajes de palabras de carácter negativo en el contexto del proyecto desarrollado en la experiencia (“diccionario ampliado”). En la tabla 2 puede observarse que esta estrategia es la que mejores resultados de clasificación de mensajes positivos y negativos ofrece, si bien sigue sin poder considerarse como buena en el caso de la identificación de los mensajes de carácter negativo.

Table 2. Resultados del conjunto de test con el análisis de sentimiento basados en diccionarios distintos.

Conjunto de entrenamiento	Método de evaluación	Clase “Positivo”	Clase “Negativo”	Total
Diccionario original	Precisión	0.947	0.123	0.900
	Cobertura	0.941	0.137	0.895
	F-Measure	0.944	0.130	0.897
Diccionario ampliado	Precisión	0.957	0.174	0.912
	Cobertura	0.906	0.329	0.873
	F-Measure	0.931	0.228	0.891

4 Conclusiones

En este trabajo se ha presentado una aplicación de técnicas de *text mining* en el ámbito educativo, con el objeto de identificar el carácter positivo o negativo de los mensajes intercambiados por los estudiantes que trabajan juntos en un equipo de proyecto.

Los análisis realizados muestran que los estudiantes emplean una terminología “pobre” en sus interacciones. Por otro lado, se observa que la mayoría de los mensajes tienen un carácter positivo, lo que, siendo bueno para el correcto desarrollo de la experiencia, se convierte en un problema a la hora de entrenar clasificadores con buenos niveles de precisión para los mensajes más negativos. Por último, se ha identificado como crucial la incorporación de términos propios del dominio de estudio en el análisis de sentimiento basado en diccionarios.

En trabajos futuros, los autores se plantean aumentar el conjunto de entrenamiento para tener una mayor población de mensajes negativos, así como aplicar otras técnicas de balanceo de datos, al mismo tiempo que emplear otras técnicas de clasificación.

Referencias

1. Alba-Eliás, F., González-Marcos, A., & Ordieres-Meré, J. (2013). An ICT based project management learning framework. In EUROCON, 2013 IEEE, pages 300–306.

2. Alba-Eliás, F., González-Marcos, A., & Ordieres-Meré, J. (2014). An active project management framework for professional skills development. *International Journal of Engineering Education*, 30(5), 1242–1253.
3. Berggren, C., Söderlund, J. (2008). Rethinking project management education: Social twists and knowledge co-production. *International Journal of Project Management*, 26, 286–296.
4. Choa, H., Gayb, G., Davidson, B., and Ingraffé, A. (2007). Social networks, communication styles, and learning performance in a cscl community. *Computers & Education*, 49(2):309–329.
5. Consejo de la Unión Europea (2004). “Education and Training 2010”. *The Success of the Lisbon Strategy Hinges on Urgent Reforms* [Online]. Council document 6905/04 of 03 march 2004. Disponible en: http://ec.europa.eu/education/policies/2010/doc/jir_council_final.pdf
6. Danneman, N. (2014). *Social media mining with R*. Packt Publishing.
7. George, S. and Leroux, P. (2002). An approach to automatic analysis of learners’ social behavior during computermediated synchronous conversations. In Cerri, S., Gouardères, G., and Paragua, F., editors, *Intelligent Tutoring Systems, volume 2363 of Lecture Notes in Computer Science*, pages 630–640. Springer Berlin Heidelberg.
8. Girasoli, A. J. and Hannafin, R. D. (2008). Using asynchronous av communication tools to increase academic self-efficacy. *Computers & Education*, 51(4):1676–1682
9. González-Marcos, A., Alba-Eliás, F., & Ordieres-Meré, J. (2015). An analytical method for measuring competence in project management. *British Journal of Educational Technology*, In Press.
10. González-Marcos, A., Alba-Eliás, F., Ordieres-Meré, J., Alfonso-Cendón, J., & Castejón-Limas, M. (2016). Learning Project Management Skills in Engineering through a Transversal Coordination Model. *International Journal of Engineering Education*, 32(2B), 894–904.
11. Hammond, M. (2005). A review of recent papers on online discussion in teaching and learning in higher education. *Journal of Asynchronous Learning Networks*, 9(3):9–23.
12. Hrastinski, S. (2008). The potential of synchronous communication to enhance participation in online discussions: A case study of two e-learning courses. *Information & Management*, 45(7):499–506.
13. Office of Government Commerce, 2009. *Managing Successful Projects with PRINCE2™*. Office Of Government Commerce
14. Ojiako, U., Ashleigh, M., Chipulu, M., Maguire, S. (2011). Learning and teaching challenges in project management. *International Journal of Project Management*, 29, 268–278.
15. Pérez-Rosas, V., Banea, C., Mihalcea, R. (2012). Learning lexicons in Spanish. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC 2012)*. European Language Resources Association (ELRA), Istanbul, Turkey.
16. Rooij, S. W. Van. (2009). Scaffolding project-based learning with the project management body of knowledge (PMBOK®). *Computers & Education*, 52(1), 210–219.
17. Russell, M. A. (2013). *Mining the social web, 2nd edition*. O’Reilly Media.
18. Thomas, J., Mengel, T. (2008). Preparing project managers to deal with complexity – Advanced project management education. *International Journal of Project Management*, 26(3), 304.
19. Turner, J.R., Keegan, A., Crawford, L. (2000). Learning by experience in the project-based organization. In: *Project management research at the turn of the millenium: proceedings of PMI research conference*, 21–24 June, Paris, France. Sylva (NC): Project Management Institute, 445–456.
20. Vapnik, V. 2008. *Statistical Learning Theory*. Wiley, Chichester, GB.