



Revista Brasileira de Pesquisa em
Turismo

E-ISSN: 1982-6125

edrbtur@gmail.com

Associação Nacional de Pesquisa e Pós-
Graduação em Turismo
Brasil

Mendes Thomaz, Guilherme; Biz, Alexandre Augusto; Michelotti Bettoni, Eduardo; Souza
Pavan, Cecilia

Modelo de monitoreo de las redes sociales para orientar en la toma de decisiones de las
Destination Management Organizations

Revista Brasileira de Pesquisa em Turismo, vol. 9, núm. 2, mayo-agosto, 2015, pp. 196-
220

Associação Nacional de Pesquisa e Pós-Graduação em Turismo
São Paulo, Brasil

Disponible en: <http://www.redalyc.org/articulo.oa?id=504151942002>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal

Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

Modelo de monitoreo de las redes sociales para orientar en la toma de decisiones de las Destination Management Organizations

Modelo de monitoramento das redes sociais para orientar a tomada de decisões das Destination Management Organizations

Social media monitoring model to guide decision making of Destination Management Organization

Guilherme Mendes Thomaz¹

Alexandre Augusto Biz²

Eduardo Michelotti Bettoni³

Cecilia Souza Pavan⁴

Resumen: La amplia adopción de los medios sociales por los usuarios y organizaciones ha generado un aumento exponencial de los datos y contenido que ofrecen oportunidades para ser tratados y transformados en informaciones y conocimientos. Sin embargo, el tratamiento con esos datos se le considera un reto y exige el uso de recursos adecuados. Considerando ese escenario, Se objetivó con ese estudio presentar un modelo de monitoreo de los medios sociales como soporte en la tomada de decisiones por las DMOs. El modelo propuesto ha sido testado en el período de pre y durante la Copa del Mundo de Fútbol FIFA 2014, utilizando las ciudades de Curitiba y Foz de Iguazú (Estado de Paraná) como destinos y, llevando una estructura de siete fases. Los resultados del período fueron compuestos por cerca de 50 mil mensajes, recorridos y asociados ontología de aplicación. Una gran parte de las relaciones de frecuencia, entre atractivos o entre ciudades anfitrionas, han sido compatibles con la realidad de su popularidad en el trade turístico.

Palabras-clave: Monitoreo de las redes sociales. Tomada de decisiones. Destination management organization.

¹ Universidade Federal do Paraná. Bolsista CNPq e Mestrando do Programa de Pós-Graduação em Turismo PPGTUR. Pesquisador do Laboratório TURITEC. E-mail: guimendesthomaz@gmail.com

² Universidade Federal do Paraná. Departamento de Turismo. Docente do Mestrado em Turismo UFPR. E-mail: biz-detur@gmail.com

³ Observatórios SESI/SENAI/IEL FIEPR Pesquisador e desenvolvimento do Observatórios SESI/SENAI/IEL. Pesquisador do Laboratório TURITEC. E-mail: du@odois.org

⁴ Universidade Federal do Paraná Bolsista CNPq e Mestranda do Programa de Pós-Graduação em Turismo PPGUT UFPR. Pesquisadora Laboratório TURITEC. E-mail: cissapavan@gmail.com

Resumo: A ampla adoção das mídias sociais pelos usuários e organizações vem gerando um aumento exponencial de dados e conteúdos que oferecem oportunidades para serem tratados e transformados em informações e conhecimentos. No entanto, o tratamento destes dados se considera um desafio e exige o uso de recursos adequados. Considerando este cenário, apresenta-se nesta pesquisa um modelo de monitoramento das redes sociais como suporte a tomada de decisões pelas *Destination Management Organizations* (DMO). O modelo proposto foi testado no período de pré e durante a Copa do Mundo de Futebol FIFA 2014, tendo como objeto os destinos de Curitiba (PR) e Foz do Iguaçu (PR) e uma estrutura de sete fases. Os resultados do período foram compostos por cerca de 50 mil mensagens, analisados e associados a uma ontologia de aplicação. Uma grande parte das relações de frequência entre atrativos turísticos e entre as cidades analisadas foram compatíveis com a realidade de sua popularidade no trade turístico.

Palavras-chave: Monitoramento de redes sociais. Tomada de decisões. Destination management organization.

Abstract: *The widespread adoption of social media by users and organizations has generated an exponential increase in data and content that offer opportunities to be processed into information and knowledge. However, dealing with such data is considered a challenge and requires the use of adequate resources. Considering this scenario, the aim of this study was to present a social media monitoring model to support decision making by DMOs. The model proposed has been tested before and during the 2014 FIFA World Cup, using Curitiba and Foz do Iguaçu (Paraná State) as destinations, having a structured of seven stages. The results of the period were composed of about 50,000 messages, tours and associated ontology application. A large part of the frequency relations, between attractions or between host cities, have been compatible with their reality in the tourist trade.*

Keywords: *Social media monitoring. Decision making. Destination management organization.*

1 INTRODUCCIÓN

La llegada de la Web 2.0 principalmente los medios de comunicación y redes sociales ocasionaron cambios y transformaciones significativas en la actividad turística, sobre todo en relación a la producción y el intercambio de informaciones y contenidos por usuarios y consumidores (UGC) ganando popularidad entre las actividades *online* de los viajeros (Xiang & Gretzel, 2010). Los contenidos publicados involucran hechos, opiniones, evaluaciones, impresiones, sentimientos, rumores y experiencias que se crean, se inician, intercambian y utilizan entre los usuarios en relación a los productos, marcas, servicios y problemas (Blackshaw & Nazzaro, 2006; Torres, 2009).

El crecimiento de las tecnologías de la información y comunicación (TIC), principalmente el desarrollo y la propagación de las medias sociales y dispositivos móviles como los smartphones y tablets contribuyen para aumento de la cantidad de datos y contenidos multimedia estructurados y no estructurado de personas, organizaciones y de los destinos turísticos, llamado de “Big Data” (Minazzi, 2015). Sin embargo, el tratamiento con esos datos se le considera un reto y exige el uso de recursos adecuados. Entre ellos están las técnicas, herramientas, procedimientos y algoritmos conocidos comúnmente como minería de datos (DM), integrantes de un macroproceso llamado descubierta de conocimiento en bancos de datos (KDD) (Santos, 2009). Al efectivo control y tratamiento de esa problemática se utiliza el término minería de contenidos en los medios sociales o *social media mining* (SMM) (Dey; Haque; Khurdiya & Shroff, 2011; Governatori & Iannella, 2011).

Por ser responsables, entre otras actividades, por el planeamiento, gestión, promoción y

desarrollo de la actividad turística en sus respectivos destinos, es fundamental que las Organizaciones de Gestión de Destinos (DMO) estén atentas en este escenario de oportunidades que el ambiente de los medios sociales ofrece. La DMO tiene como objetivo aumentar la competitividad del destino y comercializarlo a los potenciales visitantes. Sus principales responsabilidades y actividades implican la comercialización de los destinos, a coordinar el turismo, los servicios de vigilancia y estándares de calidad, proporcionan beneficios económicos y facilitar las relaciones con la comunidad local y otros grupos de interés (Pike, 2004; Blain; Levy & Ritchie, 2005). Otros investigadores, como Lau, Lee e Ho (2005), OMT (2007) e Volgger e Pechlaner (2014), también apuntan la importancia de la DMO para el desarrollo turístico.

Las informaciones y conocimientos obtenidos pueden asumir un importante rol en la gestión, en el planeamiento estratégico, desarrollo y fomento de la actividad turística, así como ofrecer oportunidades para procesos de innovación, desarrollo y la creación de nuevos productos y servicios turísticos. Considerando ese escenario, se justifica la importancia de la investigación de DMO como eje central para el desarrollo turístico de un destino.

Este estudio tuvo como objetivo presentar un modelo de monitoreo de los medios sociales como soporte en la toma de decisiones por las DMOs. El modelo propuesto ha sido testado en el período antes y durante el Mundial de Fútbol FIFA 2014 con los destinos de Curitiba y Foz de Iguaçu (Estado de Paraná). En su estructura hay siete fases. La primera fue la definición de los objetivos y el alcance del recorrido según: período, ubicación y términos de búsqueda. La segunda fue la elección del *software* de búsqueda, la tercera fase se inició con la colecta de términos de la ontología de aplicación. En la cuarta fase los resultados de todas las muestras fueron recopiladas en una sola base de datos donde se trataron todos los resultados ambiguos, inconsistentes y confusos. Ya en la quinta fase se utilizó la técnica de clúster o clústeres que permite agrupar términos (palabras-clave) y asignan significados y valores similares, la sexta etapa es el post-procesamiento, análisis, interpretación y evaluación de las normas, los problemas y las relaciones descubiertas y extraídos por las tendencias del proceso de minería de datos. Por último, la séptima fase es el procesamiento y la representación de los resultados obtenidos mediante la realización de los pasos anteriores utilizando técnicas de visualización y datos, a saber, en ese estudio, mediante un sistema de mapas heurísticos.

Para ello, fue necesario un mejor entendimiento sobre la descubierta de conocimiento, sus modelos y sus técnicas, la creación y articulación de recursos tecnológicos que permitiesen colocar el modelo en funcionamiento y aplicación de la minería de contenido en los medios sociales en un *corpus* de destaque en la actividad turística, representando los resultados de modo que facilitase la tomada de decisiones por los gestores de DMOs. Cree que este estudio pueda rellenar un espacio entre la gestión de los destinos, la evaluación de los medios sociales y el comportamiento de las personas.

2 MINERÍA DE CONTENIDOS EN MEDIOS SOCIALES Y SUS OPORTUNIDADES EN LA ACTIVIDAD TURÍSTICA

El monitoreo y análisis de contenidos en redes sociales de internet permite la detección y descubrimiento de nuevos conocimientos, identificar las opiniones, los comentarios y las evaluaciones de los consumidores y usuarios, realizar un análisis cualitativo y cuantitativo, predecir eventos y tendencias, identificar las quejas y preocupaciones de manera oportuna para evitar las crisis y el descrédito; diseñar nuevas estrategias, orientar la toma de decisiones, así como el desarrollo de innovación de productos y servicios, entre otros. (Dey, Haque, Khurdiya & Shroff, 2011; Governatori & Iannella, 2011; Paine, 2011; Zeng, Li & Duan, 2012; Han, Kamber & Pei, 2012; He, Zha & Li, 2013; Chua & Bannerjee, 2013).

El concepto de la minería de contenidos en los medios sociales (SMM) aparece como una variación de *web mining*, el proceso que involucra el desarrollo de herramientas y aplicación de técnicas para coleccionar, monitorear, analizar, resumir y visualizar contenidos de los medios sociales (Zeng, Chen, Lusch & Li, 2010; Zafarani, Abbasi & Liu, 2014). Paine (2011) defiende el monitoreo de los medios sociales como algo continuo y que demanda parámetros de comparación como: diferentes períodos, eventos, sucesos u organizaciones (concurrentes). Al relacionar los conceptos a la actividad turística, es posible pensar en monitorear la retroalimentación de los visitantes sobre sus vacaciones, identificar opiniones positivas y negativas de los consumidores sobre el destino, productos y servicios, identificar fuerzas y debilidades del plan de marketing, ayudar en la elaboración de estrategias de marketing y promoción turística, así como comprender la imagen creada y percibida del destino en la mente de los consumidores (Ku & Chen, 2007; Carson, 2008).

Las DMOs pueden inclusive utilizar las informaciones adquiridas en los medios sociales para elaborar nuevas estrategias, tomar decisiones operacionales y elaborar nuevos productos y servicios. Además, los autores creen que esta “inteligencia competitiva” puede ayudar organizaciones a identificar ventajas, debilidades, oportunidades, incentivar efectividad y mejorar la satisfacción de los consumidores (Lau, Lee e Ho, 2005). La adopción del monitoreo por esas organizaciones permitirá identificar donde están y quiénes son los consumidores, qué buscan, qué está siendo comentado sobre el destino (positiva y negativamente), entre otros. Todas estas informaciones colectadas y monitoreadas en los medios sociales se pueden utilizar estratégicamente para alcanzar los objetivos estipulados, identificar lo que está funcionando y definir las próximas acciones de marketing, planeamiento y gestión del destino turístico.

Las medias sociales también ofrecen oportunidades a DMO y OTP para reformular sus modelos de negocios, actividades, tareas, procesos y operaciones (reingeniería) bien como el desarrollo de nuevos productos y servicios, marketing, redes y gestión del conocimiento. Además, las medias sociales tienen comprobado ser una excelente estrategia para la innovación y promoción de productos, servicio y destinos turísticos (Zeng & Gerritsen, 2014; Mistilis, Buhalis & Gretzel, 2014). La innovación tecnológica es identificada como uno de los principales pilares de la ventaja competitiva para destinos turísticos, ya los gestores deben ser capaz de imaginar, percibir y eva-

luar los efectos de la ciencia y de la tecnología que se aproxima de la demanda, oferta y distribución (Dwyer, Edwards, Mistilis, Roman & Scott, 2009).

Los datos y las informaciones relevantes de los consumidores, productos, servicios y de los destinos turísticos están disponibles en Internet y en las medias sociales, pero, son poco utilizados para la optimización de los productos y servicios, bien como ofrecer soporte la decisión de los gestores de turismo. Aunque los métodos de inteligencia de los negocios y extracción del conocimiento sean puesto en muchos dominios de viajes y turismo, las aplicaciones actuales generalmente lidan con distintos procesos separadamente y carecen de un abordaje de análisis (Höepken, Fuchs, Keil & Lexhagen, 2015)

He, Zha y Li (2013) afirman que SMM se ha convertido en una necesidad crítica de las organizaciones y que tiene como objetivo la realización de un análisis más profundo y dar apoyo en las tomadas de decisiones. Para Zhang (2011), solamente las organizaciones capaces de continuamente crear y efectivamente gestionar los nuevos conocimientos sobre los clientes son capaces de adquirir ventaja e inteligencia competitiva. Aunque muchas organizaciones turísticas ya han reconocido la importancia del uso de los medios sociales como fuente de información, comunicación e interacción, aún es necesario establecer las necesidades y métodos de uso. Se destaca que el problema ni siquiera se concentra en la obtención de los datos, sino en su tratamiento de modo que sirvan a los objetivos operacionales y estratégicos (Dai, Kakkonen & Sutinen, 2011; Kavanaugh, Fox, Sheetz, Yang, Li, Shoemaker & Xie, 2012; Minazzi, 2015).

3 DESAFIOS EN EL MONITOREO DE LOS MEDIOS SOCIALES

La minería de contenidos en los medios sociales es un campo emergente donde hay más problemas que soluciones listas. Ante el gran volumen y de la naturaleza dinámica de los contenidos de los medios sociales generados continuamente, coleccionar e identificar automáticamente temas emergentes y de interés en medio a la vibración de las constantes conversiones e interacciones entre usuarios, está señalado por diversos autores como uno de los principales retos en el proceso SMM. Parte significativa del contenido se elabora ignorando reglas de ortografía y gramática; presentando problemas léxicos y sintácticos como, jergas, abreviaciones, ajustes de palabras, uso de *emotions*, creación de nuevas palabras, significados múltiples, entre otros (Paine, 2011; Abrahams, Jiao, Fan, Wang & Zhang, 2013; Zafarani, Abbasi & Liu, 2014).

Chen, Amiri, Li & Chua (2013) presentan dos retos que hay que enfrentar en el proceso de detección de temas emergentes siendo, i) el problema en coleccionar contenidos relevantes en tiempo real, ii) la capacidad de modelar las características emergentes de los tópicos detectados e identificarlos antes que se conviertan en temas populares. De acuerdo con Paine (2011) una alternativa para monitorear conversaciones en los medios sociales en tiempo real es el acompañamiento de *hashtags*⁵, una forma más universal de acompañar el contenido con referencias a los

⁵ Utilización del símbolo “#” seguido de una palabra, que se aplica cuando el usuario pretende que su publicación se

eventos o asuntos específicos. Aun así, no es la solución definitiva, teniendo en vista que los términos presentan variaciones de significado, causando recuperación de datos indeseados en medio al objeto de minería (Coutinho, Lang & Mitschang, 2013).

Paine (2011) también destaca la colecta de contenidos como la fase más desafiadora del proceso, una vez que servicios y softwares de SMM no ofrecen garantía de la integridad de los datos, pues ni todo el total de contenido será colectado y de estos ni todos serán relevantes. Por lo tanto, se comprende que lo importante no es o cuánto está siendo colectado, sino la relevancia de determinada muestra para los objetivos propuestos. A pesar de diversas herramientas y las páginas web permiten la ejecución de fases del SMM, en la mayoría de los casos, la demanda por información estratégica tiene un alto grado de especificidad, siendo que la generalización de modelos disponibles no es suficiente, exigiendo el desarrollo de sistemas personalizados. Por lo tanto, las técnicas y herramientas del SMM para colectar, compartir, investigar y visualizar datos de los medios sociales han sido ampliamente exploradas y desarrolladas (Tang & Yang, 2012; Crooks, Croitoru, Stefanidis & Radzikowski, 2013).

4 DESARROLLO DE MÉTODOS DE MONITOREO EN LOS MEDIOS SOCIALES

En el ámbito académico, varios métodos para navegar, colectar, analizar el contenido, sentimientos y tópicos en los medios sociales han sido propuestos. Muchas técnicas de categorización de datos no estructurados fueron adaptadas y aplicadas en los medios sociales en estudios de diversas áreas del conocimiento. Estos esfuerzos de investigación exigen habilidades interdisciplinarias, pues involucran desde el trato de datos brutos cuanto a la búsqueda por informaciones de calidad y con significado, exhibidas del modo más adecuado a la situación (Abrahams *et al.*, 2013).

Nikolov (2012) destaca que debido al volumen y complejidad de los datos, los modelos simples son inadecuados y, por lo tanto, el proceso de exploración de datos de los medios sociales exige un abordaje amplio, destacando la necesidad de una estructura unificada para explorar la estructura de los datos con eficiencia para realizar la identificación, clasificación y previsión de tópicos, asuntos y modelos relevantes. Innúmeras técnicas como extracción, clasificación o categorización, análisis de características, análisis lingüístico, análisis de contenido, asociación entre textos, agrupamiento y resumen se pueden usar para extraerse modelos o conocimientos interesantes e inesperados en documentos textuales (Bastos, 2006).

Para Chen & Liu (2004), la técnica de agrupamiento ofrece la ventaja de revelar tendencias imprevistas, correlaciones, o modelos en la estructura de los datos que no habían sido presupuestas. Lin, Hsieh & Chuang (2009) destacan que el análisis de agrupamiento es una técnica bien explorada en la minería de datos y, según Abdous, He & Yen (2012) esa técnica exploratoria permite visualizar modelos agrupando palabras y términos similares o también valores atribuidos que son codificados de forma semejante. Los mismos autores afirman aún que bajo una perspectiva de minería de datos, agrupamiento es la descubierta no supervisada de un modelo de datos ocultos y

que este abordaje se utiliza en situaciones en que un conjunto de registros pre-clasificados está disponible.

Las técnicas de minería de datos no aplicadas en los medios sociales para obtener conocimiento y nuevas ideas en diversas áreas como identificación de comunidades o grupo; propagación de información; la influencia de la propagación; identificación y monitoreo de tópicos y tendencias; análisis de comportamiento individual; análisis de comportamiento de grupos; detectar e identificar eventos y, claro, investigación de marketing para organizaciones (Barbier & Liu, 2011). Entre las distintas metodologías de minería de contenidos en los medios sociales destaque a Kalampokis, Tambouris & Tarabanis (2013), Abrahams *et al.*, (2013), Hea, Zha & Li (2013) y Neves (2013); al presentar: proceso de construcción metodológica, fases y etapas detalladas; diagrama de todo el proceso; aplicación en una de los medios sociales investigados en el presente estudio y; actualidad. Además de los criterios mencionados, la metodología propuesta por Neves (2013) y Hea, Zha & Li (2013) presentan aplicación directa en la actividad turística, conforme cuadro 1.

Cuadro 1- Metodologías de SMM

AUTORES	ETAPAS Y FASES
Kalampokis, Tambouris & Tarabanis (2013)	Compuesto por 2 fases: Fase de Condicionamiento de Datos (Coleta y Filtrado de Datos de Medios Sociales Brutos y Computación de Variables Predictivas); Fase de Análisis Predictiva (Creación del Modelo de Variables Predictivas y Evaluación del Desempeño Predictivo).
Abrahams <i>et al.</i> , (2013)	Selección de un fórum de discusión (1); Identificación y Colecta de Contenidos en fóruns de discusión y medios sociales (2); Selección de los términos empleados y categorías determinantes de los términos más expresivos (3); almacenamiento de los contenidos y términos seleccionados en base de datos (4); Minería de texto (5); clasificación automática de componentes (6); Lista de términos significativos por componente y categorías (7); Análisis (8).
Hea, Zha & Li (2013)	Compuesto por 3 fases: Pre-Procesamiento de Texto (Extracción, Preparación y Colecta de Contenidos); Procesamiento y Análisis (Colecta de contenido, aplicación de técnicas de minería de texto como extracción, categorización y agrupamiento, Resultados; <i>Actionable Intelligence</i> (Visualización de los resultados para identificar modelos, cuestiones, tendencias y modelos; Recomendaciones y Acciones).
Neves (2013)	Compuesto por 2 fases: Pre-Análisis (Colecta, Validación y Selección de los Contenidos); Exploración del Material (Identificación, Clasificación, Direccionamiento y Normalización de los Contenidos).

Fuente: Autores (2015)

Otros modelos identificados en la literatura también merecen destaque. El proceso de Descubierta del Conocimiento en Banco de Datos (KDD) propuesto por Fayyad, Piatetsky-Shapiro & Smyth (1996) por el pionerismo, aceptación y uso reconocido por otros investigadores. El modelo de KDD de Han, Kamber & Pei (2012) por la actualización y por ser autores clásicos en el área de minería de datos. Ya el modelo *Cross-Industry Standard Process for Data Mining* (CRISP-DM) se destaca por considerarse una de las metodologías más populares, completas y de mayor aceptación actualmente para aumentar el éxito de proyectos de minería de datos y por ser presentada en muchas publicaciones del área. Las fases y etapas de los modelos mencionados son presentadas en el Cuadro 2.

Cuadro 2- Modelos de descubierta de conocimiento en banco de datos (KDD)

NOMBRE	ETAPAS Y FASES	AUTORES
PROCESO DE KDD	Compuesto por 5 etapas: Selección dos Datos (1); Pre-procesamiento y Limpieza de Datos (2); Transformación de Datos (3); Minería de Datos (4); Interpretación y Evaluación de los Resultados (5).	Fayyad, Piatetsky-Shapiro & Smyth (1996).
PROCESO DE KDD	Compuesto por 7 etapas: Limpieza de Datos (1); Integración de Datos (2); Selección de Datos (3); Transformación de Datos (4); Minería de Datos (5); Evaluación de Modelos (6); Presentación del Conocimiento (7).	Han, Kamber & Pei (2012)
CRISP-DM	Compuesto por 6 etapas: Entendimiento del negocio (1); Entendimiento y Comprensión de los Datos (2); Preparación de los Datos (3); Modelado (4); Evaluación (5); Utilización o Aplicación (6)	Chapman, Clinton, Kerber, Khabaza, Reinartz, Shearer & Wirth (2000)

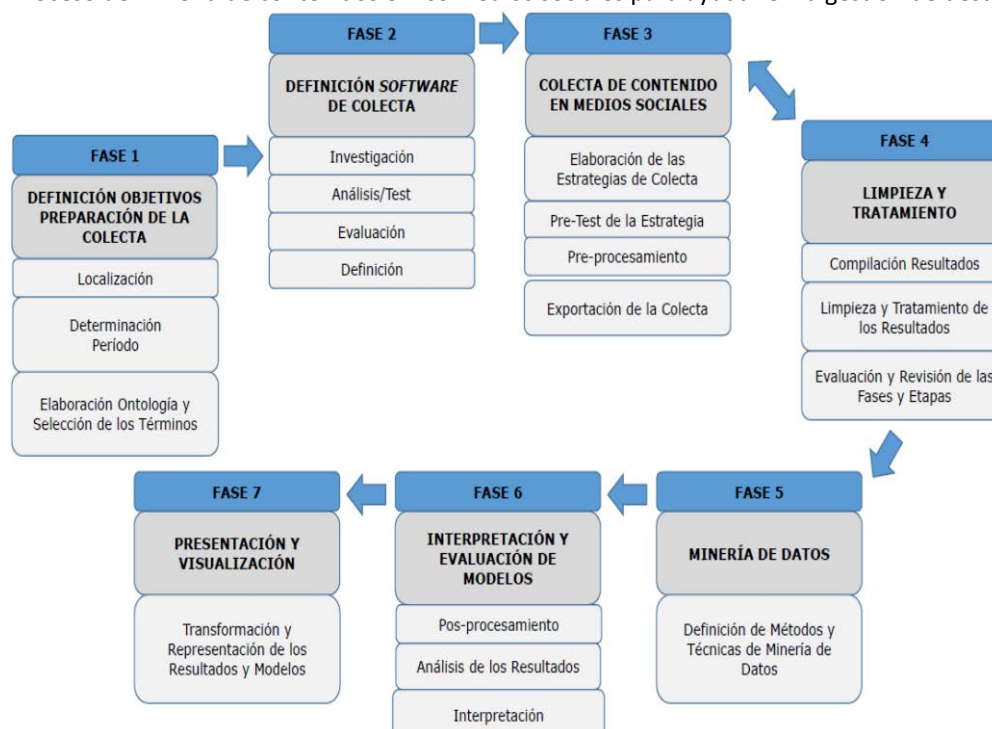
Fuente: Autores (2014)

A pesar de los innúmeros métodos y técnicas, independiente del medio de comunicación social a ser investigada, la minería de datos aplicada puede exigir abordajes y algoritmos únicos y diferentes conjuntos de datos y cuestiones requieren diferentes tipos de herramientas. El problema en sí puede determinar el mejor abordaje y después del entendimiento del problema, investigadores y analistas deben seleccionar un abordaje de minería de datos apropiada y resaltan la necesidad de la realización de la etapa de pre-procesamiento con el intuito de eliminar informaciones irrelevantes o fuera del objetivo, así como el período de tiempo a ser investigado y analizado (Barbier & Liu, 2011).

5 METODOLOGÍA

Para la construcción del proceso de minería de contenidos se utilizó como base los modelos y/o metodologías de Fayyad, Piatetsky-Shapiro y Smyth (1996), Han, Kamber y Pei (2012), Chapman *et al.* (2000), Kalampokis, Tambouris y Tarabanis (2013), Abrahams *et al.* (2013), Hea, Zha y Li (2013), y Neves (2013). Ellos tuvieron cierto éxito en el monitoreo de las redes sociales que utilizaron las siguientes técnicas, a saber: i) interpretación, anotación, clasificación, *clustering* resúmenes de los contenidos y otras técnicas de asociación y correlación.

Figura 1 - Proceso de minería de contenidos en los medios sociales para ayudar en la gestión de destinos turísticos



Fuente: Autores (2014)

El modelo propuesto empieza en la definición de los objetivos y el alcance del recorrido según: período, ubicación y términos de búsqueda. Fue testado mirando el Mundial de Fútbol FIFA 2014 y la red social escogida fue Twitter, teniendo en vista su popularidad y regularidad en el uso de la limitación de caracteres. La variable temporal, fue atendida con la definición de cinco períodos de análisis: (i) Pre-sorteo (01 set. a 25 nov. 13); (ii) Sorteo (26 nov. 13 a 31 en. 14); (iii) Pre-copa A (01 feb. a 31 mar. 14); (iv) Pre-copa B (01 abr. a 10 jun. 14) e; (v) Copa (11 jun. a 15 jul. 14).

Como el propio objetivo Copa del Mundo FIFA 2014 tenía un escopo amplio para la validación, se optó en realizar el estudio en dos ciudades del Estado de Paraná, la primera Curitiba (capital do Estado) como ciudad-sede de la Copa del Mundo FIFA 2014 y Foz de Iguazú como destino turístico internacional. Para seleccionar los términos en Twitter en inglés, portugués y español fue construida una ontología de aplicación a partir del modelo Neves (2013). Un extracto de la ontología creada se presenta en el Cuadro 3.

Cuadro 3 - Extracto de la ontología de dominio

ESTRATEGIA	DIVISIÓN	UNIDAD SEMÁNTICA	TÉRMINOS DE BÚSQUEDA
Atractivos	Curitiba	Ilha do Mel	ilha do mel, isla de la miel, isla de miel, honey island.
		Ópera de Arame	ópera de arame, opera de alambre, wire opera house
	Ciudad B	(...)	(...)
		Nombre del atractivo	término de búsqueda a, término de búsqueda b
	(...)	(...)	(...)

Fuente: Autores (2014)

La *estrategia* es un sector de interés estudiado (atractivos turísticos), la *división* distingue grupos relevantes (ciudades anfitrionas de la Copa), la *unidad semántica* (US) es el nivel de objeto significativo más específico dentro de los supuestos del estudio, mientras que los términos de búsqueda es la palabra-clave que será incluido en el sistema para que las asociaciones a una US sean encontradas.

En la segunda fase, la elección del *software* de búsqueda Seekr Monitor ocurrió en un estudio comparativo de la oferta en este segmento basado en la evaluación de las redes sociales en internet disponibles para monitoreo, las condiciones de carga de términos de búsqueda, la cantidad de mensajes, lenguas, costos del uso del *software*, formato de exportación, recuperación de los datos pasados, pre-procesamiento de los datos. La tercera fase se inició con la colecta de términos de la ontología de dominio de las dos ciudades-objetivo, tanto para atractivos como para servicios, en el período citado anteriormente. En esa misma fase hubo un constante acompañamiento de resultados parciales para captar alguna eventual anomalía en los modelos de frecuencia de los datos. Así, algunas inconsistencias ya fueron tratadas antes de la fase 4.

En la cuarta fase de los resultados de todas las muestras fueron recopiladas en una sola base de datos donde se trataron todos los resultados ambiguos, inconsistentes y confusos, y las estrategias se trabajaron juntos en una base de datos MariaDB para solucionar problemas donde se establecieron los datos. Dos métodos principales fueron aplicados: (a) análisis de frases más frecuentes y (b) análisis de los termos más frecuentes y sus correlaciones. En lo método (a) fue posible identificar y eliminar tendencias que no estaban relacionadas al objetivo de la coleta pero causaban parcialidad, ejemplo, ambigüedad entre los nombres de los atractivos de los destinos con homónimos de otros Países. Ya lo método (b) fue posible validar o refutar tópicos de la ontología propuesta, además, identificar temas emergentes no previstos inicialmente, aspectos esenciales citados en parte significativa de la literatura de referencia. Fueran creados reglas de tratamiento y a partir de esto, permitió reducir los ruidos de los datos inadecuados.

En la quinta fase fue utilizada una minearía de contenido automatizada, hechas por la clasificación automática de ocurrencias por medio de la existencia de termos (palabras-clave) y sus significados. Los casos pudieran ser codificados de acuerdo con su unidad semántica (US) y división de acuerdo con la ontología. La sexta etapa es el post-procesamiento, análisis, interpretación y evaluación de las normas, los problemas y las relaciones descubiertas y extraídos por las tendencias del proceso de minería de contenido. En lo caso del objeto de estudio, eso se dio por la formulación de tablas de frecuencia de divisiones, US y palabras-clave para cada elemento de la ontología y de acuerdo con las respectiva fase del evento (seria histórica). En esas tablas fueran aplicados color rojo de las más frecuentes para los menos frecuentes, facilitando su inspección visual por las posibles anomalías no previstas. La actitud de esa esta fue reforzada cuando de la identificación de algunas oblicuidades, siendo nuevamente evaluados y los resultados en correcciones, y los casos irrelevantes fueran sacados. Por último, la séptima fase es el procesamiento y la representación de

los resultados obtenidos mediante la realización de los pasos anteriores utilizando técnicas de visualización y datos, a saber, en ese estudio, mediante un sistema de mapas heurísticos⁶.

Las categorías de análisis, en función de las ontologías orientadoras, presentan una serie de relaciones de pertenencia y frecuencia. Por ejemplo: un atractivo pertenece a una ciudad-sede, que pertenece a una división – atractivos. Además, cada atractivo o tipo de servicio posee una cantidad de ocurrencias. Al cruzar las instalaciones y conteo, se optó por expresar esa relación con círculos circunscritos, cuyos tamaños representan la densidad de ocurrencias. El sistema de mapas heurísticos puede ser visto en el tópico de análisis e, inclusive, comparado al formato tabular. Lo permite una verificación rápida de los movimientos de ascensión y queda de determinados atractivos/servicios por apelo visual, característica menos evidentes en las tablas tradicionales.

6 RESULTADOS DE CURITIBA

El estudio avanzó en el sentido de extrapolar el uso de las *hashtags* evento, profundizándose en cualquier mención que haya involucrado el nombre de las ciudades y algunos atractivos y/o servicios, en los períodos de análisis. Con ello, fue posible desarrollar un trabajo más detallado, aplicándose la ontología de aplicación para identificar elementos constituyentes de la actividad turística que tuvieron oscilaciones de popularidad. En ese detalle, fue posible analizar muestras de mensajes con mayor ocurrencia, validando el método aplicado.

La llamada cuarta fase del modelo, específicamente para el destino turístico de Curitiba, fue generada la tabla 1, después de los tratamientos necesarios de los datos no sobresalientes. Los datos no deseados fueron sacados, pues es fundamental para el suceso de lo SMM. En la misma dirección, validar los datos recogidos corrobora con el modelo propuesto. Al enumerar los mensajes más frecuentes, no se buscó definir las más importantes o influyentes, en virtud de la variabilidad prevista. El objetivo, mismo en un alcance pequeño, fue validar el método de obtención de los datos y encontrar inconsistencias. Los padrones pueden se repetir en parte de la base, mismo en frases no exactamente iguales.

⁶ Con base en la biblioteca D3JS en JavaScript (D3 DATA-DRIVEN DOCUMENTS, 2014)

Tabla 1 - Mensajes con más frecuencia en el destino curitiba

Mensaje	Ctd.	% del destino
[suprimido] Concejal: JARDIM BOTÂNICO CONHEÇA CURITIBA http://t.co/7ejodsKgGR	58	0,44%
[suprimido] Concejal: PARQUE TANGUA CONHEÇA CURITIBA http://t.co/2BozT9zwq4	57	0,43%
[suprimido] Concejal: MUSEU OSCAR NIEMEYER CONHEÇA CURITIBA http://t.co/riO1OPv38A	75	0,57%
RT @CUNALi_YILDIZ: Frost in Botanical Garden of Curitiba (the coldest capital city of Brazil) http://t.co/enfejZ0AbD	18	0,14%
RT @BpMarcelloFJU: En visita al Jardín Botánico de Curitiba, saqué esta foto de un Quiero-Quiero, en homenaje a os VPR\'s do Brasil. http://t.co...	17	0,13%
NX Cero es una de las atracciones de la Fifa Fan Fest. La banda se presenta en Curitiba, en el Parque Barigui. http://t.co/bRzYx6574j	16	0,12%
RT @ZagonelMarcelo: Trabajo de hoy para la Feria Brasil Moda Show! De 01* a 11 de Mayo en el Parque Barigui. Curitiba-PR. http://t.co/MDGoFvYsWK	15	0,11%
RT @CUNALi_YILDIZ: The Wire Opera House in Curitiba, Southern Brazil http://t.co/oFGAS7zEli	13	0,10%
RT @gazetadopovo: Restaurante de Curitiba crea postre en homenaje al Museo Oscar Niemeyer. http://t.co/8F5WSyBPpi http://t.co/MYrRyccv...	12	0,09%
RT @CUNALi_YILDIZ: Good night at Botanical Garden of Curitiba, Southern Brazil! http://t.co/gILGb2F2Js	11	0,08%
Total		2,31%

Fuente: Autores (2014)

Los resultados de la fase de post-procesamiento (fase sexta) para el destino Curitiba están relacionados en la tabla 2. Los atractivos relacionados en la ontología correspondiente, fueron investigados con el fin de medir su popularidad del período de interés de la investigación. En lo recorte son presentados los diez principales atractivos del destino.

Tabla 2 - Ocurrencias de atractivos - Curitiba

Atractivo	Atractivos – Ocurrencias (Destino Curitiba)					Total
	Pre-sorteo	Sorteo	Pre-copa A	Pre-copa B	Copa	
Jardim Botânico	1138	1499	996	723	326	4.682
Parque Barigui	503	671	441	533	163	2.311
MON - Museo Oscar Niemeyer	409	546	286	409	217	1.867
Barrio de Santa Felicidade	285	330	317	411	134	1.477
Sector Histórico de Curitiba	218	247	194	221	205	1.085
Parque Tanguá	185	318	120	27	35	685
Ópera de Arame	186	203	67	74	11	541
Torre Panoâmica	76	96	64	61	15	312
Ilha do Mel	18	50	22	32	18	140
Estrada da Graciosa	3	3	24	4	9	43
Total	3.021	3.963	2.531	2.495	1.133	13.143

Fuente: Autores (2014)

Con base en la tabla 2 es posible verificar que tanto la frecuencia de la presencia de los atractivos reconocidos en los mensajes, cuanto una concentración de menciones en fases antes del Mundial de Fútbol 2014. Como soporte al gestor de DMO, hay indicios de que lo movimiento

pre-evento fue más importante do que lo evento y, sería fundamental ter trabajados con esos atractivos durante el período evaluado.

En el caso de la aplicación del análisis de servicios, vale destacar que por una cuestión de la propia madurez del método en relación a los atractivos, el período denominado “Pre-Sorteo” no llegó a ser computado. En la época, tal ontología aún estaba en fase de experimento y no se pudo aprovechar dado el posible sesgo con relación a los datos. De la misma forma que para los atractivos, los datos tienen dos formas de presentación. La table 3 traz los datos de ocurrencia de servicios en Curitiba.

En lugar del ocurrido con los atractivos (tabla 2), los servicios (tablas 3) hubo concentración mayor de mensajes en lo período anterior al inicio del evento. Para un DMO, eso puede ser un direccionamiento de donde invertir recursos de la promoción y atención al turista de acuerdo con la proximidad del evento: necesidad esta que puede ser verificada por medio de lo SMM.

El formato tabular presentado es el más tradicional para la representación de frecuencia de ocurrencias, principalmente cuando ya existe un escopo de categorías Pre-existentes. Sin embargo, dado el contexto orgánico y dinámico del contenido encontrado en los medios sociales, la emergencia de categorías/temas, además de las varias relaciones de pertenencia entre categorías, formas alternativas de representación de los movimientos en internet podrían favorecer a los gestores, como es el caso de los mapas heurísticos. La estrategia de servicios se representó en las siguientes páginas.

Tabla 3 - Ocurrencias de servicios - Curitiba

Servicio	Servicios – Ocurrencias (Destino Curitiba)					Total
	Pre-Sorteo	Sorteo	Pre-copa A	Pre-copa B	Copa	
Aeropuerto	-	1726	2732	4290	1382	10.130
Hotel	-	1029	1320	1721	901	4.971
Restaurante	-	965	995	1636	363	3.959
Churrasco	-	1086	863	1077	282	3.308
Bebida	-	292	184	360	160	996
Tren	-	163	177	348	62	750
Bicicleta	-	161	241	209	61	672
Bar	-	100	86	82	47	315
Hostal	-	104	55	88	44	291
Taxi	-	229	140	98	39	506
Cena	-	177	83	116	33	409
Almuerzo	-	209	148	55	24	436
Sightseeing	-	102	23	24	19	168
Desayuno	-	61	44	18	10	133
Posada	-	4	4	5	0	13
Acomodaciones	-	3	5	1	0	9
Ómnibus	-	3	0	2	0	5
Alquiler de Carros	-	0	0	0	0	0
Total	-	6.414	7.100	10.130	3.427	27.071

Fuente: Autores (2014)

Vale señalar un detalle en cuanto a las figuras 2: el presente formato – artículo – causa algunas limitaciones en la presentación del sistema de mapas heurísticos, por la necesidad de ser una publicación estática. Así, se hicieron capturas de la pantalla del sistema que, en su funcionamiento, presenta interactividad para que el usuario pueda aproximarse o apartarse de determinada categoría del dominio, además de alterar algunos datos. Así, la herramienta adquiere características más dinámicas y evita que las mismas informaciones tengan que ser presentadas en informes extensos a los gestores.

Como forma de facilitar el entendimiento, se toma como ejemplo el caso del *Sightseeing*. Orientándose inicialmente por la tabla 3, se nota que el tópico se inició con destaque y después tuvo una caída que lo acompañó prácticamente hasta el último período. Tal relación también se nota fácilmente en la columna “Transportes”, dado el tamaño del círculo con subtítulo *Sightseeing* a lo largo del tiempo.

Figura 2 - Destino Curitiba

(continua)

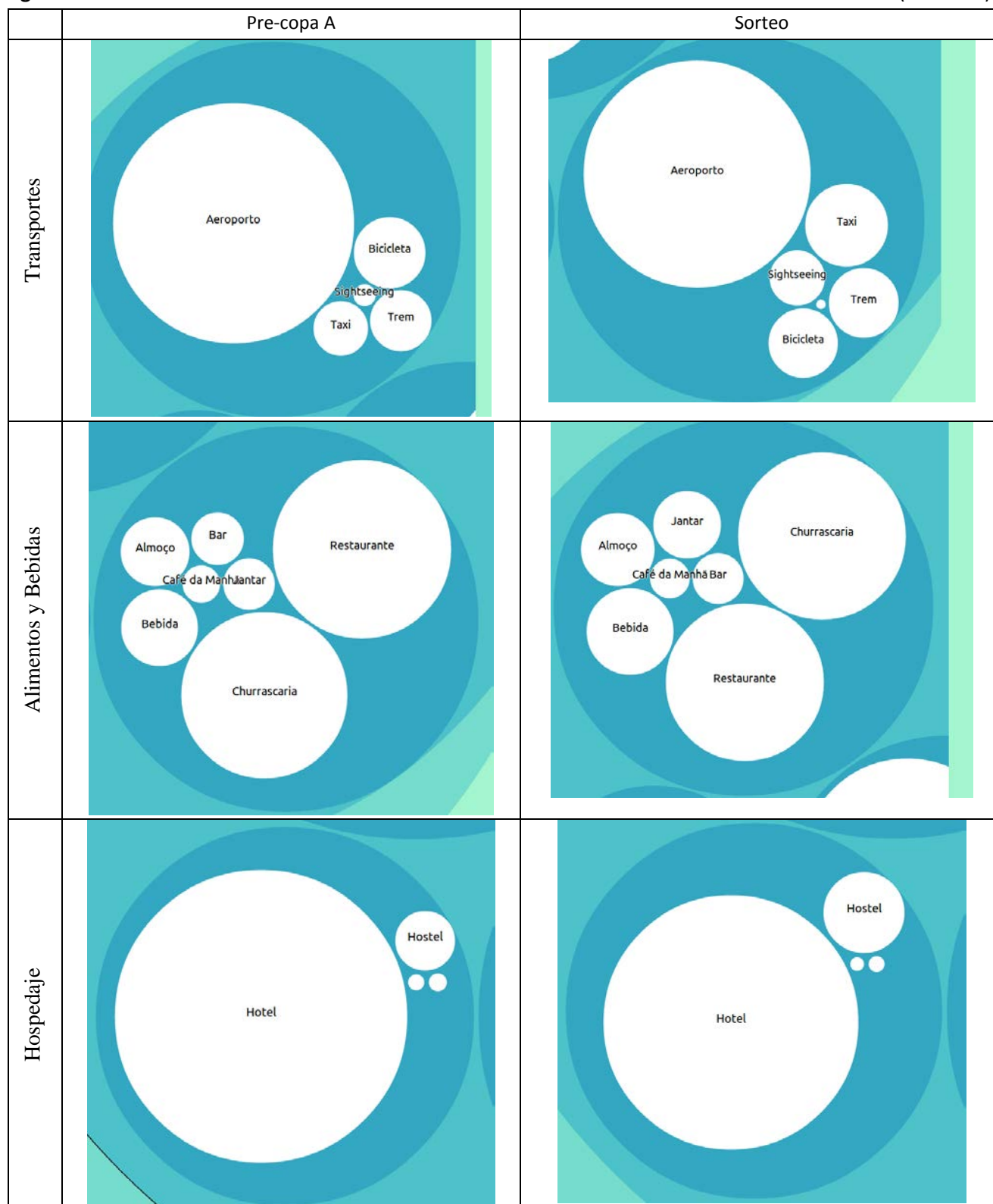
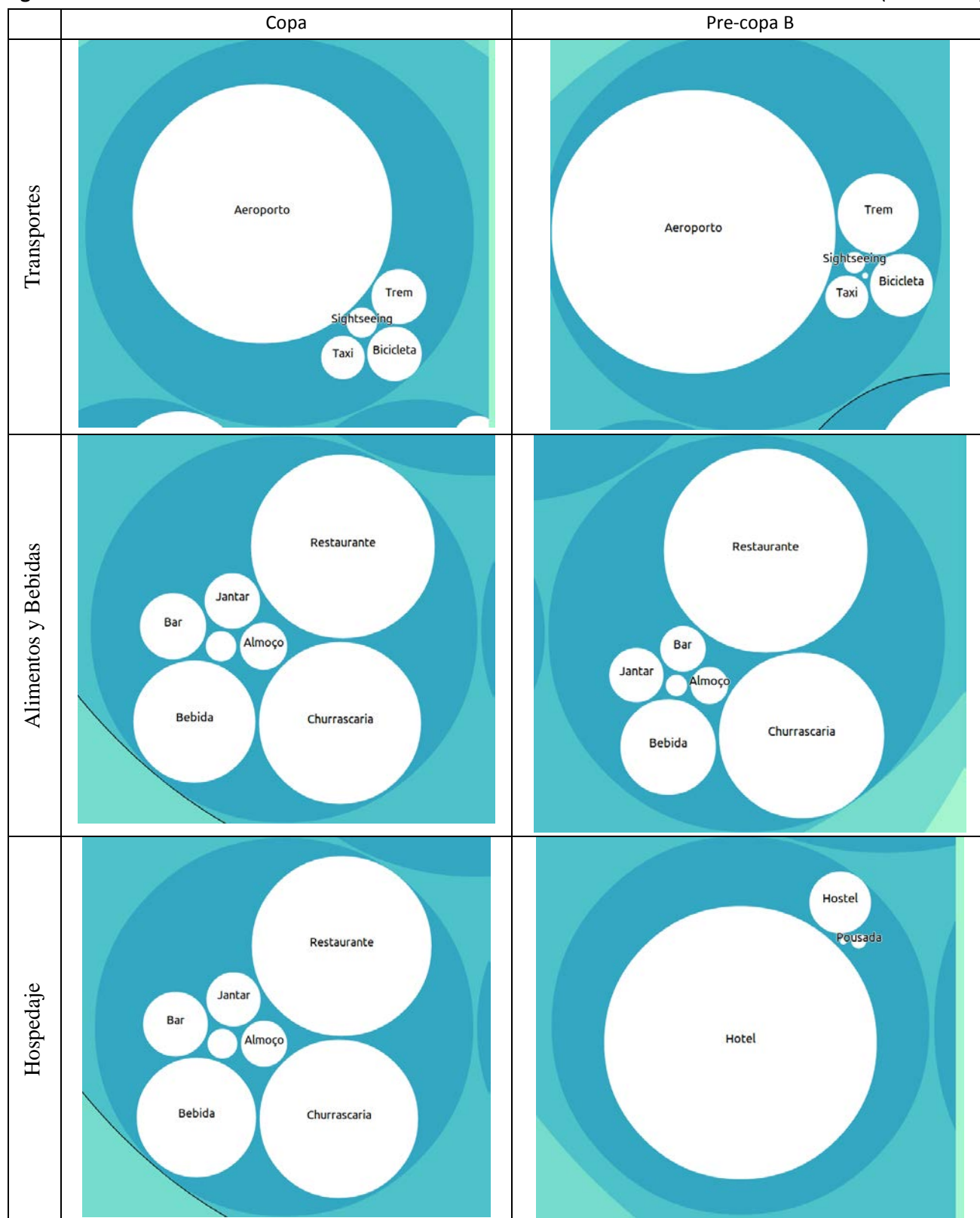


Figura 2 - Destino Curitiba

(Conclusão)



Fuente: Autores (2014)

Los datos representados por medio de mapas heurísticos son los mismos de formato tabular. Las conclusiones cuanto al crecimiento de tópicos y/o su importancia también son similares, pero, con un mayor apelo visual y posibilidades de navegación y acompañamiento de las fases, sien necesariamente quedarse por las cuantías específicas de cada uno. A ejemplo, al presentar gráficamente que “bebida” y “restaurante” pasan a ganar relevancia y “café da manhã” pierde su expresión con lo pasar de las fases, estrategias específicas de la promoción y su relación con lo trade turístico pueden ser ejecutadas de acuerdo con las expectativas de DMO para aquel tópico. Esta reacción y la tomada de decisión con base en la opinión del usuario son los beneficios apuntados anteriormente y en lo cual la función tornarse más esencial con o pasar de los años.

7 RESULTADOS DE FOZ DE IGUAZÚ

La misma técnica se utilizó en Foz de Iguazú, donde el intercambio fue más frecuente y el asunto principal fue el Aeropuerto de la ciudad, con 165 replicaciones de la misma noticia sobre las obras de mejoras y manutención, conforme Tabla 4.

Tabla 4 - Mensajes con más Frecuencia en el destino FOZ De Iguazú

Mensaje	Ctd.	% del destino
[AEROPUERTOS] En Foz do Iguazú, Aeroporto Internacional Cataratas está con la pista cerrada, hasta las 12h00, para obras de mejoras.	165	1,44%
[TRÁNSITO] Boletín con el tráfico en el Puente de la Amistad y en las principales vías de circulación en Foz de Iguazú. Oiga en vivo en CBN Foz.	129	1,13%
RT @ElUniversal: #NoDejeDeVer Turistas visitan las Cataratas del Iguazú en Foz de Iguazú, en Paraná. (AFP) http://t.co/QQ7l6p1Jy4	32	0,28%
RT @Kefera: NAMO AMITUOFO 🙌 @ Templo Budista 🙏 Foz do Iguazú http://t.co/pibCZr2C9u	32	0,28%
RT @AvellarPaulo: Cataratas de Foz do Iguazú, Brasil ... http://t.co/TveGITvW6h	31	0,27%
RT @Kefera: ☺ @ Mezquita Árabe 🕌 Foz do Iguazú http://t.co/6Bf7H92K75	24	0,21%
RT @Kefera: 📍 @ Templo Budista 🙏 Foz do Iguazú http://t.co/L9SOctdoAI	20	0,17%
RT @guia4rodas: Cataratas do Iguazú, queridinhas brasileiras (y argentinas también): http://t.co/lyOprtCpFo http://t.co/m7DUef3R7R	18	0,16%
RT @cbnfoz: [TRÁNSITO] Boletín con el tráfico en el Puente de la Amistad y en las principales vías de circulación en Foz de Iguazú. Oiga en vivo en CBN...	17	0,15%
RT @hoteismabu: Vean qué guapos los nuevos vecinos del @ParqueDasAves en Foz de Iguazú. #MuitoBemVindo #HoteisMabu http://t.co/OGg7DgTANF	13	0,11%
Total		4,21%

Fuente: Autores (2014)

Del mismo modo que en Curitiba, solamente los mensajes más frecuentes no pudieron responder por más del 5% del total. Aun así, se notó la reincidencia de mensajes replicados a partir de una personalidad que visitó el destino, a *vlogger* Kéfera, popular entre adolescentes. Para el destino Foz de Iguazú también se consolidó una tabla de frecuencia de atractivos según la ontología, conforme Tabla 5.

Tabla 5 - Ocurrencias de atractivos – Foz de Iguazú

Atractivo	Atractivos – Ocurrencias (Foz de Iguazú)					Total
	Pre-Sorteo	Sorteo	Pre-copa A	Pre-copa B	Copa	
Cataratas do Iguazú	1307	1711	1657	1959	692	7.326
Usina Hidroelétrica de Itaipu	343	337	292	493	164	1.629
Parque Das Aves	102	153	104	183	46	588
Rafain Churrascaria Show	153	155	85	100	48	541
Puente de la Amistad	30	81	196	114	17	438
Parque Nacional de Iguazú	58	121	61	110	11	361
Templo Budista	124	32	25	29	30	240
Mezquita Musulmana	97	16	12	25	10	160
Marco das Tres Fronteiras	15	20	24	56	17	132
Casino Iguazú	2	16	0	1	1	20
Total	2.231	2.642	2.456	3.070	1.036	11.435

Fuente: Autores (2014)

Las ocurrencias de los atractivos corroboran los supuestos en cuanto a la popularidad y frecuencia de las visitas en el destino Foz de Iguazú, lo que ayuda en la legitimación del método y de la ontología de dominio. En el caso de Foz de Iguazú, la tendencia fue semejante en relación a los servicios principales, como se puede observar en la Tabla 6.

Tabla 6 - Ocurrencias de Servicios - Foz de Iguazú

Servicio	Servicios – Ocurrencias (Destino Foz de Iguazú)					Total
	Pre-Sorteo	Sorteo	Pre-copa A	Pre-copa B	Copa	
Aeropuerto	-	502	669	882	300	2.353
Hotel	-	656	708	753	210	2.327
Churrasco	-	128	105	108	56	397
Restaurante	-	83	87	144	42	356
Hostal	-	44	94	105	33	276
Bebida	-	33	67	63	19	182
Cena	-	14	8	12	9	43
Taxi	-	2	5	6	9	888
Posada	-	12	21	30	6	69
Desayuno	-	14	19	17	6	56
Bicicleta	-	7	10	3	4	24
Bar	-	9	2	3	1	15
Almuerzo	-	9	4	7	1	21
Acomodaciones	-	0	1	0	0	1
Ómnibus	-	0	0	0	0	0
Tren	-	0	0	1	0	1
Alquiler de Carros	-	12	3	1	0	16
Sightseeing	-	0	5	0	0	5
Total	-	1.525	1.808	3.001	696	7.030

Fuente: Autores (2014)

El sistema del mapa heurístico también fue preparado para los resultados del destino Foz de Iguazú (Figura 3). Al contrario del caso de Curitiba, donde algunos elementos tuvieron una

variación no proporcional, en el caso de Foz de Iguazú hubo una tendencia general mediana de los servicios en aumento hasta las proximidades del evento para iniciar una caída después (Hotel, Aeropuerto, Restaurante, Hostal). Teniendo en vista ese hecho, el mapa heurístico va a mantener la misma forma en todo período pues éste se construyó en base a la proporción entre los elementos. Si todos los servicios suben proporcionalmente, el “más frecuente continua siendo el más frecuente” y ocupa un círculo mayor. El mapa sólo se va alterar cuando algún movimiento desproporcional suceda, por ejemplo, finalización de la obra del Aeropuerto cuando supuestamente la situación va a estabilizarse y los avisos cesarán.

Figura 3 - Destino Foz de Iguaçu (PR)

(continua)

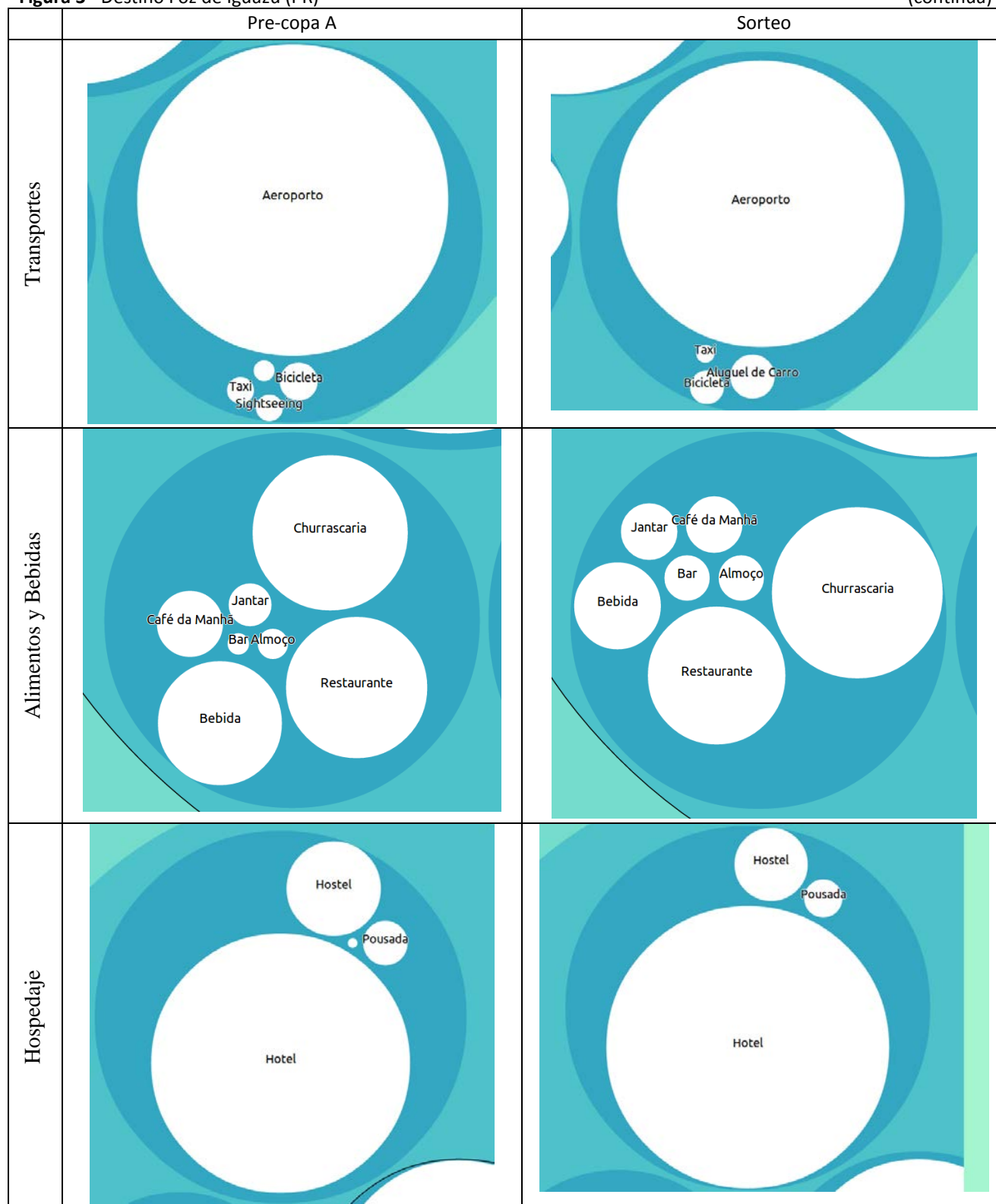
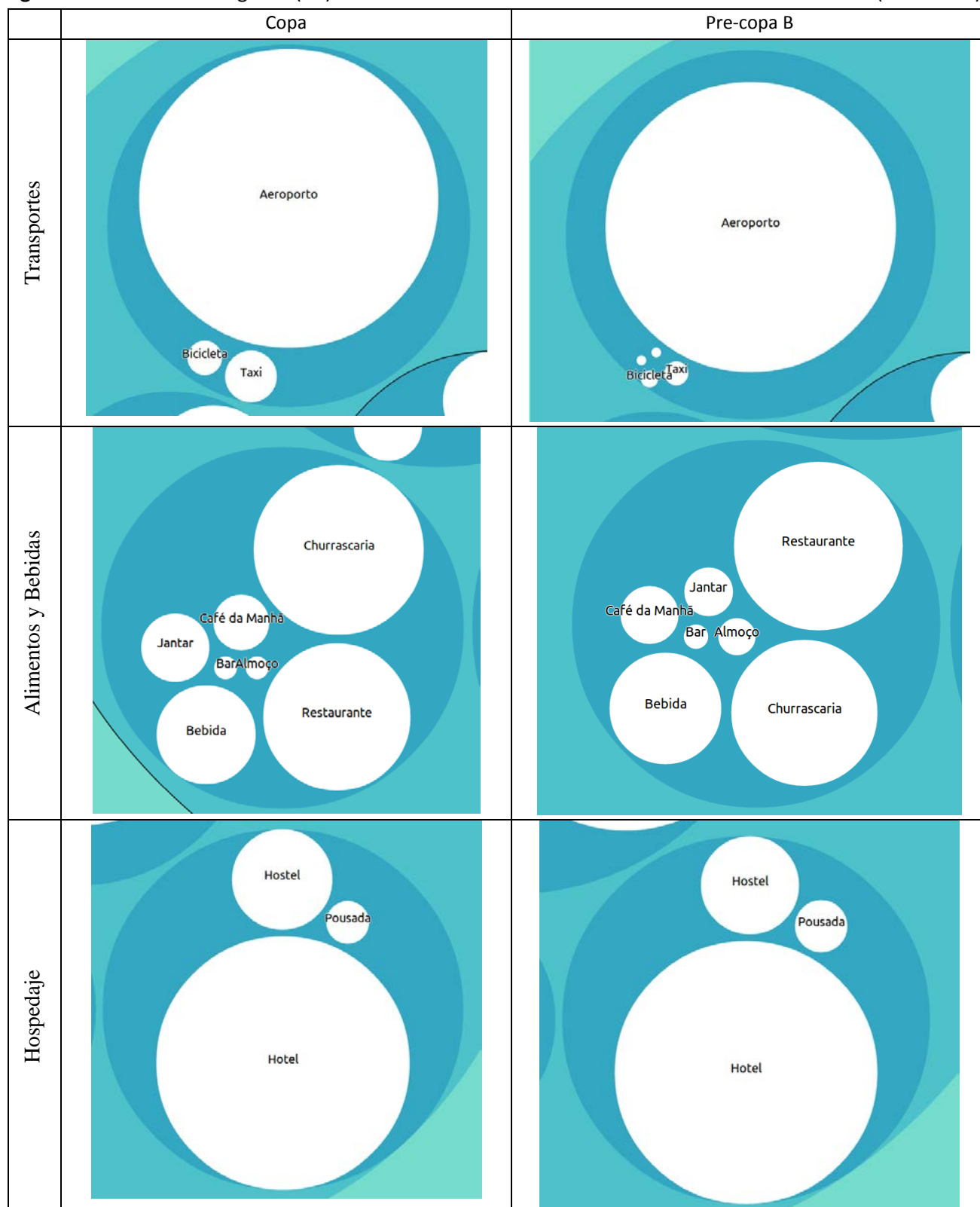


Figura 3 - Destino Foz de Iguazú (PR)

(Conclusão)



Fuente: Autores (2014)

8 CONCLUSIONES

Es importante destacar que a cada fase de la investigación el tipo de contenido publicado presentó características distintas como, por ejemplo, antes del Sorteo de la ciudad-sede y de las selecciones la búsqueda era por informaciones más genéricas, a lo contrario de la Pre-copa donde la búsqueda se concentró más en atractivos, medios de transportes, otros. En este sentido, se confirma la importancia del monitoreo en distintas fases del evento acompañando las necesidades de los turistas y al mismo tiempo oportunidades de negocios.

La gran cuestión del monitoreo está en la construcción ontológica y consecutivamente de las semánticas entre las palabras-clave. Como discutido en procedimientos metodológicos los resultados de los mensajes con más frecuencia no representan los más importantes e influenciados, teniendo en vista el alto grado de variabilidad encontrado en la base. Mas, pueden dar importantes indicios de las tendencias o mismo de la cualidad de los datos que están siendo recogidos. Por lo tanto, el análisis cualitativa de los contenidos tiene un importante peso en el monitoreo de los medios sociales.

Los resultados del período fueron compuestos por cerca de 50 mil mensajes, recorridos y asociados ontología de aplicación – la suma de las dos ciudades-objetivo. Una gran parte de las relaciones de frecuencia, entre atractivos o entre ciudades anfitrionas, han sido compatibles con la realidad de popularidad de estas US's en el *trade* turístico. Esto refuerza la legitimidad de este método propuesto, además de componer un primer diagnóstico, a ser comparado con otros en una serie temporal, permitiendo que las oscilaciones sean identificadas y estudiadas con el fin de apoyar los procesos en la tomada de decisiones en las DMO's.

En comparación a los modelos y métodos presentados, el estudio difiere por una combinación de las fases y pasos, así como para permitir el desarrollo de diversas estrategias para el recogido de datos de forma automática a través del *software* de monitoreo de redes sociales, lo que facilita y hace el proceso más eficaz al enfocarse en contenidos más relevantes. Además, no se limita a un escopo de tiempo, red social u ontología, teniendo en vista que permite adaptaciones y aplicaciones en otras áreas y contextos, con las adaptaciones de un dominio que se convierte en el objeto. Los resultados del estudio pueden apoyar tanto las *Destination Management Organizations* (DMO) como a las organizaciones turísticas en la tomada de decisiones más enérgicas que se basan en el descubrimiento de conocimiento.

Aún que dos destinos tengan sido utilizados como teste para lo modelo, no fue objetivo de este trabajo describir y analizar en profundidad la fluctuación y la importancia de los tópicos durante lo evento en cuestión. En la investigación se buscó presentar y evaluar uno modelo de monitoreo con datos reales, con lo cual las relaciones de frecuencia al longo del tiempo pueden ser recogidas, tratadas y presentadas para que, al final, potenciales gestores de DMO puedan tomar decisiones al relacionar datos y conocimientos previos.

Presentan como factores limitantes de la investigación la ontología de aplicación de servicios pues los termos aún son limitados que necesitan ampliar para contemplar todas las

posibilidades y servicios turísticos, y la gran cantidad de publicaciones colectadas y lo software utilizado para o monitoreo de las medias sociales no posibilitaran los análisis de todo contenido. Como sugerencia para investigaciones futuras es la ampliación de las ontologías de aplicación elaboradas y para monitoreo y minería de contenidos en medias sociales para otras DMO mirando los Juegos Olímpicos de Rio en 2016.

REFERENCIAS

- Abrahams, A. S., Jiao, J., Fan, W., Wang, G. A., & Zhang, Z. (2013) What's buzzing in the blizzard of buzz? Automotive component isolation in social media postings. *Decision Support Systems*, 55 (4), 871-882.
- Abdous, M., He, W.; & Yen, C. (2012). Using data mining for predicting relationships between online question theme and final grade. *Educational Technology & Society*, 15 (3), 77-88.
- Barbier, G., & Liu, H. (2011). *Data mining in social media*, in Aggarwal, Charu C. (ed) Social Network Data Analytics, Springer US, 327-352.
- Bastos, V. M. (2006). *Ambiente de Descoberta do Conhecimento na Web para Língua Portuguesa*. Tese (CO-PPE). Universidade Federal do Rio de Janeiro (UFRJ): Rio de Janeiro.
- Blackshaw, P., & Nazzaro, M. (2006). *Consumer-generated media (CGM) 101: Word-of-mouth in the age of the web-fortified consumer*, A Nielsen BuzzMetrics White Paper, Spring, New York.
- Blain, C.; Levy, S. E.; Ritchie, JR B. (2005) Destination branding: Insights and practices from destination management organizations. *Journal of Travel Research*, 43, (4), p. 328-338.
- Carson, D. (2008). The Blogshare as a Market Research Tool for Tourism Destination: A Case Study of Australia's Northern Territory, *Journal of Vacation Marketing*, 14 (2), 111-119.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *Crisp-Dm 1.0 - Step-By-Step Data Mining Guide*. 2000. Spss Inc.
- Chen, Y.; Amiri, H.; Li, Z., & Chua, T. S. (2013) Emerging topic detection for organizations from microblogs. In: *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2013. 43-52. Disponível em: <<http://dl.acm.org/citation.cfm?id=2484057>>. Acesso: 02/10/2013.
- Chen, S. Y., & Liu, X. (2004). The Contribution Of Data Mining To Information Science. *Journal of Information Science*, 30 (6), 550-558.
- Chua, A. Y., & Banerjee, S. (2013). Customer knowledge management via social media: the case of Starbucks. *Journal of Knowledge Management*, 17(2), 237-249.
- Coutinho, F. C.; LANG, A.; & MITSCHANG, B. (2013) Making Social Media Analysis more efficient through Taxonomy Supported Concept Suggestion. In: *15th GI-Symposium Database Systems for Business, Technology and Web*, 11th March - 15th March, 2013. 457-476. Disponível em: <<http://www.btw-2013.de/index-en.html>>. Acesso em: 06/10/2013.

- Crooks, A., Croitoru, A., Stefanidis, A., & Radzikowski, J. (2013). *#Earthquake*: Twitter as a distributed sensor system, *Transactions in GIS*, 17(1), 124–147.
- Dai, Y., Kakkonen, T., & Sutinen, E. (2011) MinEDec: a decision-support model that combines text-mining technologies with two competitive intelligence analysis methods, *International Journal of Computer Information Systems and Industrial Management Applications*, 3, 165-173.
- Dey, L., Haque, S. M., Khurdiya, A., & Shroff, G. (2011). Acquiring competitive intelligence from social media. In: *Proceedings of The 2011 Joint Workshop on Multilingual OCR and Analytics for Noisy Unstructured Text Data*. ACM. pp. 3.
- Dwyer, L., Edwards, D., Mistilis, N., Roman, C., & Scott, N. (2009). Destination and enterprise management for a tourism future. *Tourism Management*, 30(1), 63-74.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17 (3), 37-54.
- Governatori, G.; & Iannella, R. (2011). A modeling and reasoning framework for social networks policies. *Enterprise Information Systems*, 1 (5), 145–167.
- Han, J., Kamber, M., & Pei, J. (2012) Data mining: concepts and techniques. Waltham: Elsevier.
- He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management*, 33 (3), 464-472.
- Höpken, W., Fuchs, M., Keil, D., & Lexhagen, M. (2015). Business intelligence for cross-process knowledge extraction at tourism destinations. *Information Technology & Tourism*, 1-30.
- Kalampokis, E., Tambouris, E., & Tarabanis, K. (2013). Understanding the Predictive Power of Social Media. *Internet Research*, 23 (5), 544-559.
- Kavanaugh, A. L.; Fox, E. A.; Sheetz, S. D.; Yang, S.; Li, L. T.; Shoemaker, D. J.; & Xie, L. Social media use by government: From the routine to the critical. *Government Information Quarterly*, 29(4), 480-491, 2012.
- Ku, L.-W.; & Chen, H.-H. (2007). Mining opinions from the Web: Beyond relevance retrieval. *Journal of the American Society for Information Science and Technology*, 58(12), 1838-1850.
- Kwak, H.; Lee, C.; Park H.; & Moon S. (2010). What is Twitter, a social network or a news media? In: *Proceedings of the 19th international conference on World Wide Web*. ACM, 2010. 591-600.
- Lau, K. N., Lee, K. H., & Ho, Y. (2005). Text mining for the hotel industry, *Cornell Hotel and Restaurant Administration Quarterly*, 46(3), 344-362.
- Lin, F.-R., Hsieh, L.-S., & Chuang, F.-T. (2009) Discovering genres of online discussion threads via text mining. *Computers & Education*, 52 (2), 481-495.
- Minazzi, R. (2015). Social Media Metrics and Analysis. In: *Social Media Marketing in Tourism and Hospitality* (pp. 137-163). Springer International Publishing.
- Mistilis, N., Buhalis, D., & Gretzel, U. (2014). Future eDestination Marketing Perspective of an Australian

Tourism Stakeholder Network. *Journal of Travel Research*, 53(6), 778-790.

Nikolov, S. (2012). *Trend or No Trend: A Novel Nonparonetric Method for Classifying Time Series*, Massachusetts Institute of Technology, Cambridge, Estados Unidos.

Neves, A. J. W. A. (2013). *Qualidade percebida dos produtos e serviços turísticos em eventos: das ferramentas de análises de conteúdo do twitter*. Dissertação (Programa de Pós-Graduação em Ciência, Gestão e Tecnologia da Informação). Universidade Federal do Paraná (UFPR): Curitiba.

OMT – Organização Mundial de Turismo. (2007). *A Practical Guide to Tourism Destination Management*. Disponível em http://www.visitkerteminde.dk/sites/default/files/asp/visitkerteminde/KUP/a_practical_guide_to_tourism_destination_management.pdf > Acesso em 14 fev. 2015.

Paine, K. D. (2011). *Measure what matters: Online tools for understanding customers, social media, engagement, and key relationships*. John Wiley & Sons.

Pike, S. (2004). *Destination Marketing Organisations*. North Holland: Elsevier.

Santos, R. (2009). Conceitos de Mineração de dados na web. *XV Simpósio Brasileiro de Sistemas de Multimídia e Web, VI Simpósio Brasileiro de Sistemas Colaborativos-Anais*. MM Teixeira, CAC Teixeira, FAM Trinta, e P. PM Farias, Eds, 81-124, 2009.

Tang, X., & Yang, C. C. (2012). Social network integration and analysis using a generalization and probabilistic approach for privacy preservation, *Security Informatics*, 1(1), 1-14.

Torres, C. (2009). *Tudo o que você queria saber sobre marketing e publicidade na Internet e não tinha a quem perguntar*, Novatec, São Paulo.

Volgger, M.; Pechlaner, H. (2014). Requirements for destination management organizations in destination governance: understanding DMO success. *Tourism Management*, 41, (1), p.64-75.

Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search, *Tourism management*, 31(2), 179-188.

Zafarani, R., Abbasi, M. A., & Liu, H. (2014). *Social Media Mining: An Introduction*. Cambridge University Press.

Zeng, B., & Gerritsen, R. (2014). What do we know about social media in tourism? A review. *Tourism Management Perspectives*, 10, 27-36.

Zeng, L., Li, L., & Duan, L. (2012). Business intelligence in enterprise computing environment. *Information Technology & Management*, 4 (13), 297–310.

Zhang, Z. J. (2011). Customer knowledge management and the strategies of social software. *Business Process Management Journal*, 17 (1), 82-106.

Artigo recebido em: 24/11/2014.

Artigo aprovado em: 13/08/2015.