GALiWeatherApp: applicación móvil para predicción meteorológica individualizada en lenguaje natural

M. Fresquet-Rius, A. Ramos-Soto, A. Bugarín, and S. Barro

Centro de Investigación en Tecnoloxías da Información (CiTIUS), Universidade de Santiago de Compostela
{marc.fresquet@rai.usc.es, alejandro.ramos, alberto.bugarin.diz, senen.barro}@usc.es

Resumen En este trabajo presentamos GALiWeatherApp, aplicación Android que genera automáticamente predicciones meteorológicas en lenguaje natural a partir de los datos de predicción numérica y operativa de la Agencia Meteorológica de Galicia (Meteogalicia). Utilizando técnicas de soft computing y generación de lenguaje natural (NLG) GALiWeateherApp construye una narrativa en texto que describe el pronóstico meteorológico a 3-4 días vista, de forma individualizada y automática para cualquiera de los ayuntamientos de Galicia o para cualquier punto del mapa. Esta aplicación se ha presentado a la 'Competición de Apps basadas en técnicas de IA' organizada en la XVI Conferencia de la Asociación Española para la Inteligencia Artificial (CAEPIA 2015).

Keywords: Descripciones lingüísticas de datos, Generación de Lenguaje Natural, Sistemas 'Data to text', predicción meteorológica

1. Breve descripción de la app

GALiWeatherApp genera automáticamente en lenguaje natural la predicción numérico-simbólica que emite diariamente la Agencia Meteorológica de Galicia (Meteogalicia). Utilizando técnicas de soft computing (conjuntos borrosos) y generación de lenguaje natural (metodología 'data-to-text'), GALiWeatherApp construye una narrativa en texto que describe el pronóstico meteorológico a 3-4 días vista, de forma individualizada y automática para cualquier localización. GALiWeatherApp toma como entrada datos simbólicos y numéricos procedentes de las predicciones operativa y numérica de Meteogalicia para una localización determinada (tanto desde el listado de ayuntamientos como desde cualquier punto del mapa) y, a través del servicio Web GALiWeather [1], genera automáticamente y visualiza el comentario en texto acerca de dicho pronóstico. En dicho comentario se incluye información sobre las variables meteorológicas de interés en la predicción meteorológica operativa: cobertura nubosa, precipitación, viento (dirección e intensidad) y temperaturas (máximas y mínimas). GALiWeather modela y gestiona la incertidumbre existente en los términos de la predicción mediante conjuntos borrosos, que se utilizan de dos formas:

- para el tratamiento de referencias temporales imprecisas en la ventana temporal de 3-4 días
- para flexibilizar la exigencia del cumplimiento de una determinada condición para la variable 'cobertura nubosa' dentro de dicha ventana

Con esto, se consigue construir un mayor número de posibles descripciones lingüísticas de los eventos relevantes, que son potenciales candidatas para ser tratadas a continuación con la metodología de uso más habitual [2] en la Generación de Lenguaje Natural (NLG). Las descripciones lingüísticas corresponden a la fase inicial de dicha metodología (denominada 'determinación del contenido') y son tratadas posteriormente en las fases posteriores (planificación del documento y realización lingüística), para generar narrativas con una calidad lingüística suficiente que puede ser consumida por los usuarios de la aplicación.

2. Link del vídeo y la app

La aplicación se ha desarrollado para la plataforma Android, está plenamente funcional y puede descargarse desde la página de la aplicación [3], desde cuya URL puede accederse también al vídeo demostrativo de las funcionalidades incluidas en la app y a un breve resumen descriptivo de su funcionamiento.

3. Problema que resuelve

En la actualidad el equipo de predicción operativa de Meteogalicia elabora y publica manualmente una predicción genérica a tres días vista para toda Galicia, que se publica de forma gráfica (mediante mapas) acompañada de un comentario general en texto. Adicionalmente, se produce también una predicción a 3-4 días vista individualizada para cada ayuntamiento, que se publica mediante símbolos gráficos (denominados meteoros), como los que se muestran en la Figura 1.

Figura 1. Ejemplo de predicción operativa tomado de la Web de Meteogalicia [4], correspondiente al 16 septiembre de 2015, donde se muestran los meteoros que representan gráficamente la información relevante del pronóstico a cuatro días vista.

La elaboración de un comentario en texto individualizado para cada uno de los 314 ayuntamientos donde se recojan los eventos más relevantes del pronóstico resulta inviable por la excesiva carga de trabajo adicional que supondría para los

expertos meteorólogos y por ello la predicción meteorológica por ayuntamientos únicamente se sirve en formato simbólico y numérico. Sin embargo, la demanda de predicciones comentadas en formato texto por parte de los usuarios de información meteorológica es creciente e incluso ha sido objetivada en estudios en este ámbito [5]. Como respuesta a esta demanda nuestro equipo desarrolló conjuntamente con Meteogalicia el servicio GALiWeather para la generación automática de predicciones textuales individualizadas para cada uno de los ayuntamientos de Galicia. Una vez integrado y validado, GALiWeather fue puesto en marcha en mayo de 2015 como un servicio público plenamente funcional a disposición de los visitantes de la Web de Meteogalicia [4]. Un ejemplo de predicción textual generada automáticamente es la que mostramos en la Figura 2. Hasta donde conocemos [6], únicamente existe en el mundo otro servicio con un grado de madurez similar [5], si bien sigue una aproximación NLG tradicional, sin gestión específica de la incertidumbre de los términos.

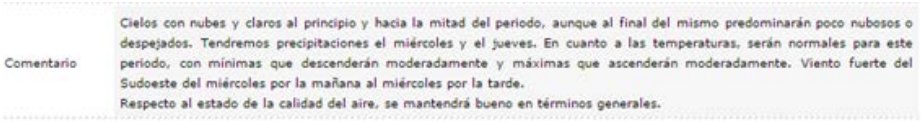

Figura 2. Ejemplo de narrativa generada automáticamente para la descripción textual de la predicción operativa de la Figura 1, tomado de la Web de Meteogalicia [4].

El servicio GALiWeather utiliza los datos simbólico-numéricos de predicción operativa, junto con información adicional de los expertos meteorólogos, para generar automáticamente narrativas que describen en forma textual la información más relevante del pronóstico individualizado por ayuntamientos. La aproximación que sigue GALiWeather integra técnicas de 'soft computing' (conjuntos borrosos) para la gestión de la incertidumbre en algunos de los términos de las narrativas, junto con estrategias convencionales de los sistemas 'Data To Text' para la generación final de los textos de predicción.

Los datos de entrada se refieren a una ventana temporal de cuatro días, con datos para cada una de las variables de interés en la predicción operativa:

- Estado del cielo: se representa por tres valores diarios (mañana, tarde y noche), mediante 21 símbolos correspondientes a dos variables diferenciadas (variable nubosa y precipitación)
- Viento: se representa por tres valores diarios (mañana, tarde y noche), mediante 34 símbolos correspondientes a la dirección e intensidad del viento
- Temperatura: se representa por un valor diario, mediante dos valores numéricos correspondientes a la temperatura máxima y mínima previstas para cada día, junto con información sobre la diferencia entre estas temperaturas y la media climática de los últimos treinta años.

A través de la nueva aplicación móvil GALiWeatherApp se ha ampliado la funcionalidad de GALiWeather y se mejora la accesibilidad a dicho servicio, ya que se ha incluido en una misma aplicación:

- la predicción operativa general, mostrada en formato gráfico simbólico mediante mapas y meteoros, como los que se muestran en la Figura 3
- se muestra también la predicción particularizada por ayuntamientos, tanto en formato simbólico como textual (generada automáticamente por el servicio GALiWeather), algo que resulta especialmente útil en dispositivos con pantallas de reducidas dimensiones, donde la visualización gráfica de la información puede presentar dificultades

Como ejemplo, mostramos en la Figura 4 el pronóstico real generado para Santiago de Compostela, correspondiente al 14 de octubre de 2015.

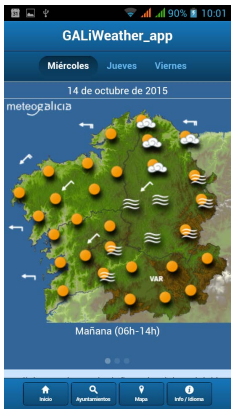

Figura 3. Captura de pantalla de GALiWeatherApp de la predicción operativa simbólica general para toda Galicia.

Adicionalmente GALiWeatherApp extiende la funcionalidad de GALiWeather, puesto que, además de incluir facilidades de búsqueda en la lista de ayuntamientos disponibles (Figura 4) se ha incluido la generación de pronósticos textuales para cualquier localización del mapa (Figura 5.

4. Técnicas de Inteligencia Artificial empleadas

En esta sección realizamos un breve resumen de las técnicas de Inteligencia Artificial utilizadas en el desarrollo del servicio GALiWeather, cuya descripción detallada puede encontrarse en [1]. Como ya se ha comentado en la sección anterior, GALiWeather extrae la información relevante de los datos simbólico-numéricos del pronóstico operativo de Meteogalicia combinando técnicas de 'soft computing' y D2T. Ello se realiza en dos etapas diferenciadas:

Figura 4. Captura de pantalla de GALiWeatherApp de la predicción operativa textual y simbólica particularizada para un ayuntamiento (pronóstico para Santiago de Compostela, 14 de octubre de 2015).

- Generación de descripciones lingüísticas, que abstraen los datos de entrada en etiquetas borrosas (valores lingüísticos y referencias temporales).
- Generación de lenguaje Natural, que convierte las descripciones anteriores en narrativas finales.

La primera etapa, cuyo esquema general mostramos en la Figura 6, obtiene para cada variable de interés un conjunto de etiquetas lingüísticas y referencias temporales que contienen la información más relevante de los datos.

El tratamiento borroso de los datos se centra en la variable 'cobertura nubosa' para proporcionar, de acuerdo con el estilo y los requisitos fijados por el meteorólogo, una descripción cronológica del comportamiento de dicha variable en la ventana temporal de 3-4 días que se considera en los pronósticos. Para ello se utiliza una partición temporal borrosa de tres etiquetas 'inicio', 'mitad'. 'final', referidas a dicha ventana temporal (Figura 7a). Cuando no resulta posible construir la descripción cronológica, se genera una descripción global a corto plazo de la variable. Para ello se utiliza una partición de tres cuantificadores borrosos 'ocasional', 'relevante', 'predominante' (Figura 7b), con los que se evalúan las proposiciones cuantificadas borrosas de tipo I formadas por dichos cuantificadores y los valores de cobertura nubosa considerados ('muy nuboso', 'parcialmente nuboso', 'claro'). Dicha evaluación utiliza el modelo de cuantificación de Zadeh [7] para la evaluación de sentencias cuantificadas, tal y como se muestra en la Figura 8.

Por su parte, las descripciones correspondientes a las variables 'precipitación' y 'viento' siguen una estrategia similar, en la que se extraen los episodios relevantes y se clasifican de acuerdo con los tipos definidos para cada variable por los meteorólogos, mediante métodos 'crisp.' Finalmente, las variables de temperatura (máxima y mínima diaria) se generan descripciones relativas a su tendencia frente a los valores climáticos (media de referencia de los últimos treinta años).

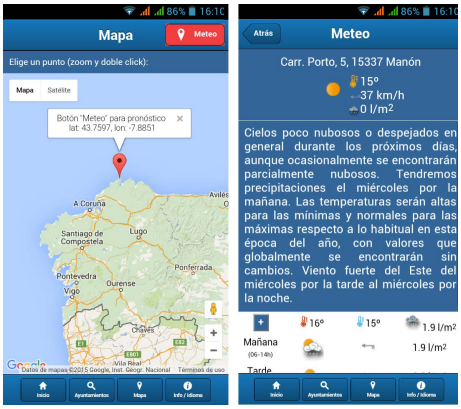

Figura 5. Acceso desde GALiWeatherApp a las predicciones para cualquier localización del mapa, junto con el pronóstico correspondiente, generado por el servicio GALiWeather (pronóstico para Estaca de Bares, 14 de octubre de 2015).

A estas descripciones lingüísticas se les aplica seguidamente la metodología estándar NLG [2]para dar lugar a las narrativas finales, mediante dos aproximaciones:

- Para el caso de frases de menor complejidad sintáctica (las relativas a las temperaturas o la cobertura nubosa) se siguió una aproximación simple basada en plantillas XML
- Para el caso de expresiones más complejas, como las relativas a la precipitación -donde, por ejemplo, puede ocurrir más de un episodio significativo-, se siguió una aproximación NLG para evitar reiteraciones, mejorar la calidad lingüística de las expresiones y conseguir que fuesen perfectamente legibles e indistinguibles de las que podría construir un experto humano

Para todo ello, fue necesario implementar en GALiWeather cuatro de las seis fases habituales en un sistema NLG:

- Determinación de contenido: decisión de qué información se comunicará en el texto final, mediante la creación de un conjunto de objetos que contienen los datos filtrados y resumidos, esencialmente las descripciones lingüísticas mencionadas anteriormente
- Agregación de sentencias: agrupación de los mensajes en sentencias, por episodios, días y agregación completa final
- Lexicalización: decisión de qué palabras y frases específicas se van a escoger para expresar conceptos y relaciones en los mensajes
- Realización lingüística, que produce un texto final correcto desde los puntos de vista sintáctico, morfológico y ortográfico

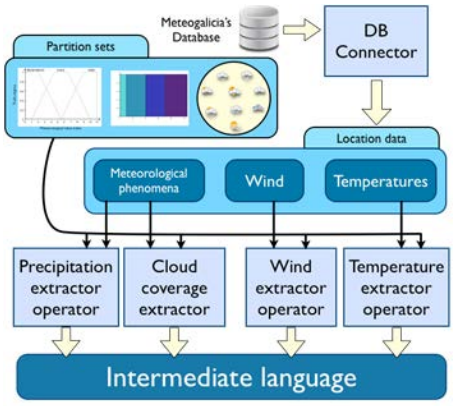

Figura 6. Esquema global del método de generación de las descripciones lingüísticas.

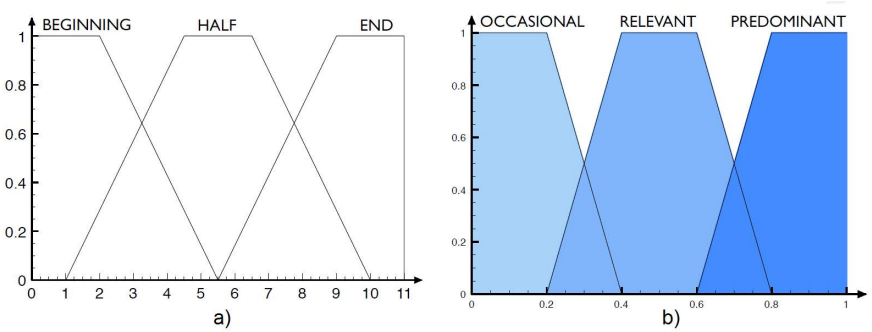

Figura 7. Particiones correspondientes a los operadores borrosos de la variable 'cobertura nubosa'.

5. Utilidad y viabilidad de la aplicación

La utilidad de la aplicación es máxima y su viabilidad total, puesto que proporciona una nueva modalidad de acceso a un servicio que en la actualidad recibe un número ingente de consultas. Baste tener en cuenta que en 2014 la web de Meteogalicia registró más de 50 millones de accesos. El hecho de que GALiWeather se haya incluido como un servicio público en la web de Meteogalicia es un indicador inequívoco del respaldo oficial que ha obtenido y del interés de este tipo de servicios para los usuarios finales. Asimismo, los resultados de evaluación de la calidad de contenido y forma de las narrativas, realizada por los expertos meteorólogos (puntuación de 4.83 sobre 5 [1]), junto con el funcionamiento ininterrumpido por varios meses del servicio, sin que se hayan producido casos de malfuncionamiento o errores detectados por los usuarios, avalan la calidad final del desarrollo.

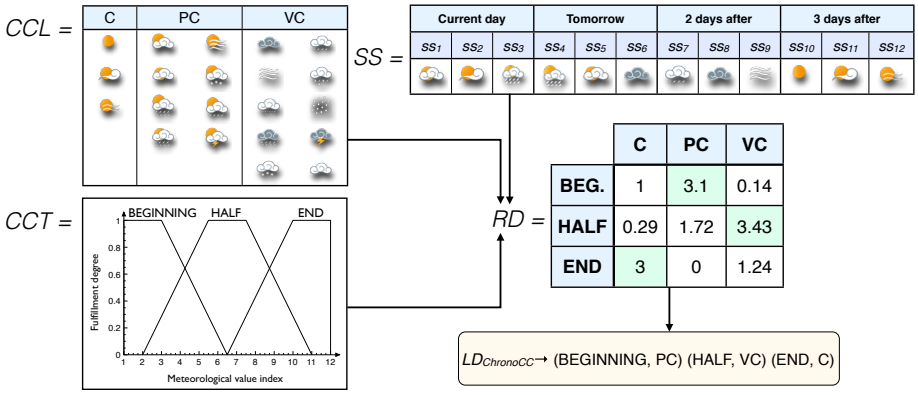

Figura 8. Evaluación de las sentencias cuantificadas correspondientes a la descripción cronológica de la variable 'cobertura nubosa'.

6. Conclusiones

Con la aplicación móvil GALiWeatherApp se ha desarrollado un caso real de aplicación de métodos y tecnologías de soft computing y NLG para la generación automática de predicciones meteorológicas en formato textual. Este tipo de narrativas comunican de forma más directa a los usuarios aquellos aspectos más relevantes de la predicción, cosa que -como avalan algunos estudios recientes [5]- no es siempre posible con las representaciones gráficas de la información. En concreto, el servicio Web GALiWeather, que da soporte a la generación de textos de GALiWeatherApp, es un ejemplo real de cómo a partir de una investigación básica en el ámbito del soft computing ha dado lugar a un servicio público plenamente disponible para los usuarios de Meteogalicia.

En futuras ampliaciones de la actual versión de GALiWeatherApp nos planteamos el añadir un sistema de síntesis de voz, que permita la narración oral de los pronósticos textuales, así como la mejora en algunos aspectos funcionales (listado de localizaciones emblemáticas o de especial interés) y técnicos (estudios de la conversión de datos de predicción numérica a operativa). Una vez realizadas estas mejoras la aplicación será puesta a disposición del público en general.

Agradecimientos Esta investigación ha sido financiada por el Ministerio de Economía y Competitividad (proyecto TIN2014-56633-C3-1-R, cofinanciado por el Programa FEDER) y por la Xunta de Galicia (proyectos EM2014/012 y CN2012/151, cofinanciados por el Programa FEDER). A. Ramos-Soto está financiado por el Ministerio de Economía y Competitividad mediante el programa de contratos predoctorales FPI (BES-2012-051878).

Referencias

1. Ramos-Soto, A., Bugarín, A., Barro, S., Taboada, J.: Linguistic descriptions for automatic generation of textual short-term weather forecasts on real prediction data. IEEE Transactions on Fuzzy Systems **23**(1) (2015) 44–57
2. Reiter, E., Dale, R.: Building Natural Language Generation Systems. Cambridge University Press (2000)
3. Fresquet-Rius, M.: GALiWeatherApp, http://proxectos.citius.usc.es/galiweatherapp/ (Last visited October15th 2015)
4. Meteogalicia: Web de predicción por ayuntamientos, http://www.meteogalicia.es/ (Last visited October15th 2015)
5. Sripada, S.G., Burnett, N., Turner, R., Mastin, J., Evans, D.: A case study: NLG meeting weather industry demand for quality and quantity of textual weather forecasts. In: INLG-2014 Proceedings. (June 2014)
6. Ramos-Soto, A., Bugarín, A., Barro, S.: On the role of linguistic descriptions of data in the building of natural language generation systems. Fuzzy Sets and Systems (2015)
7. Zadeh, L.A.: A computational theory of dispositions. International Journal of Intelligent Systems **2**(1) (1987) 39–63