

iSAC: atención ciudadana virtual con reconocimiento del lenguaje coloquial

Miquel Montaner Rigall¹, Antonio Manuel López Arjona², Josep Lluís de la Rosa Esteva² y
M^a Mercè Rovira Regàs²

(miquel@isac.cat, antonio@isac.cat, pepluis@isac.cat, merce@isac.cat)

¹ *Strategic Attention Management, S.L.*

² *Centro EASY de la Red IT del CIDEM- Universidad de Girona*

***Abstract.** Cuando nos proponemos llevar la atención ciudadana a los canales virtuales (p.ej. Internet, WAP o TDT) nos aparecen una serie de problemas técnicos que debemos afrontar. El más importante de ellos es que una máquina pueda entender las preguntas que realizan los ciudadanos con lenguaje coloquial y no con lenguaje administrativo, sobretodo teniendo en cuenta que muchos ciudadanos no saben ni expresarse ni escribir correctamente. Además, es importante resaltar que en un sistema de atención ciudadana un ciudadano realiza una pregunta y espera una sola respuesta, no miles como se encuentran en un buscador clásico. En este artículo se muestra como el sistema iSAC resuelve este problema con algoritmos de tratamiento de lenguaje natural y como consigue la concreción si inundar a los ciudadanos con información irrelevante. Así pues, iSAC permite a los ciudadanos solucionar sus dudas utilizando el lenguaje que les resulta más cómodo sin que esto afecte a la calidad del servicio dado. En los resultados experimentales se demuestra que el sistema iSAC es mucho más efectivo para la atención ciudadana que no un buscador cualquiera. iSAC es una herramienta de código abierto y está actualmente implantado en el Ayuntamiento de Terrassa dónde en 10 meses ha recibido más de 62.000 preguntas de sus ciudadanos.

***Keywords.** Atención ciudadana, canales virtuales, lenguaje natural, innovación tecnológica.

Introducción

Los sistemas de atención ciudadana de cualquier administración pública¹ sufren de forma continua de falta de recursos, este hecho dificulta el buen desarrollo de las tareas para las cuales fueron concebidos. La razón de la falta de recursos es la creciente demanda [1] que hay por parte de los ciudadanos de hacer uso de los servicios que se ofrecen.

Actualmente el servicio de atención ciudadana tiene dos canales de respuesta: uno es la atención presencial y el otro la atención por vía telefónica. Este hecho condiciona el uso que se pueda hacer del servicio a la disponibilidad horaria del personal que participa y a los

¹ <http://www.terrassa.cat/ayuntamiento/010>

recursos humanos que se destinen. Se han propuesto varias opciones por solucionar este problema:

Una opción son las centralitas telefónicas preprogramadas. Esta solución presenta deficiencias, dado el gran volumen de información que pueden contener las bases de datos del servicio de atención ciudadana, en el caso de Terrassa² más de 18.000 respuestas posibles, es difícil su implantación y mantenimiento.

Otra opción es la de ofrecer un servicio Web que implemente un motor de búsqueda estándar. Esta solución presenta problemas como el carácter exhaustivo de los motores de búsqueda, que choca con el fin de concreción de los servicios de atención ciudadana, como los problemas legales relacionados con el acceso a la información, teniendo en cuenta la *Ley Orgánica de Protección de Datos* [2], problemas con la imparcialidad que ha de asegurar una entidad pública, mirando de no favorecer nunca ningún comercio en contra de otro y finalmente limitaciones referidas al reconocimiento del lenguaje natural en que se expresan los ciudadanos.

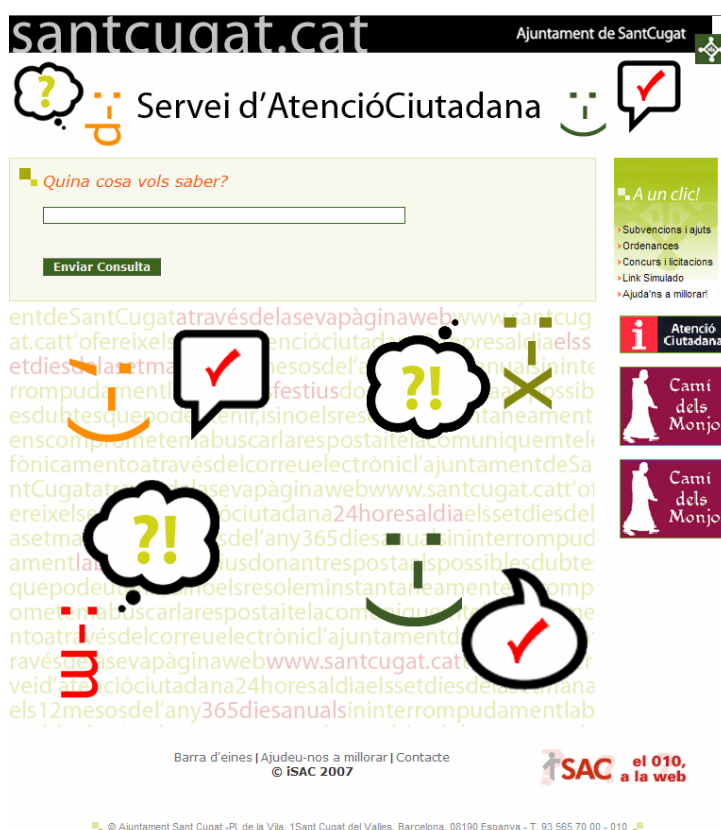
La última opción es una mezcla de las dos anteriores, un sistema web basado en FAQs. Estos sistemas están contruidos a partir de un conjunto reducido de respuestas que se desea atender. El sistema se configura con la finalidad de conseguir relacionar las posibles preguntas con el conjunto de respuestas preestablecido. Como en el primer caso, presenta problemas en cuanto al mantenimiento, puesto que añadir una nueva respuesta implica estudiar todos los casos que la deben proporcionar. También presenta problemas de escalabilidad, haciendo inviable la instalación de una herramienta de este tipo a sistemas con un conjunto elevado de respuestas. Aun cuando soluciona otros problemas cómo el del lenguaje natural, la imparcialidad o el cumplimiento de la LOPD.

Una vez presentadas las diferentes opciones, nuestra propuesta es crear iSAC: una herramienta multicanal (Web, WAP, TDT,...) que solucione los problemas de los buscadores estándares. Concretamente se quiere crear un buscador que ofrezca concreción, teniendo en cuenta que el sistema óptimo encontraría sólo una respuesta por cada pregunta, mejore aspectos relacionados con el lenguaje natural, corrija el funcionamiento para los casos que afecten a temas como la LOPD o el de la imparcialidad, ofrezca un sistema con gran facilidad de uso, fácilmente escalable y el mantenimiento del cual se haga de forma automática, con la consecuente vigencia de los datos.

Inicialmente presentamos iSAC como un buscador (véase la Figura), pero el objetivo final es que tenga una imagen de bot (agente virtual) que dialogue con los ciudadanos, es decir, que le puedan hacer preguntas y responda con una solución o con una pregunta de aclaración.

² <http://isac.terrassa.org>

Figura 1. Aspecto de iSAC en Sant Cugat del Vallès (Barcelona)



A continuación presentamos los trabajos previos que se han realizado en este campo. En la sección 2 veremos la descripción detallada de la solución propuesta. En la sección 3 se ven las diferentes pruebas de funcionamiento y las ideas que se extraen. Y finalmente, concluimos el artículo y presentamos posibles trabajos futuros.

1. Investigación previa

Hoy en día el entorno Web ha crecido en gran medida [3], esto ha dado paso a la aparición de nuevas herramientas para cubrir una necesidad hasta ahora poco relevante o inexistente

[4]. Gracias a esta nueva necesidad se ha podido profundizar mucho en técnicas de recuperación de información.

Cuando se habla de buscadores o de recuperación de información es casi inevitable no pensar en Google. Este buscador usa técnicas basadas en la democracia de la red [5]. Para hacerlo, valora para todas las páginas qué otras páginas le hacen referencia y de este modo calculan lo que ellos denominan *PageRank*, usando la cantidad de páginas que hagan referencia y a la vez el *PageRank* que tengan estas. El uso de esta herramienta es simple y transparente para el usuario. Pero presenta problemas si se tienen en cuenta los objetivos marcados en cuando a la concreción de la respuesta, el procesamiento del lenguaje natural o la imparcialidad respecto la recuperación de la información.

Otra clase de buscadores Web, como Clusty³ y Vivísimo [6], ambos procedentes de la Carnegie Mellon University, como el buscador Web Lycos, quieren solucionar los problemas del exceso de información devuelta aplicando técnicas de Clustering [7], pero no acaban de resolver el tema de la exhaustividad, sólo dan la herramienta al usuario para que este mismo haga la elección de resultados, además, siguen sufriendo el resto de problemas de Google en cuanto a la LOPD y la imparcialidad.

Otros sistemas [8] hacen uso de análisis semántico y mejoran en general el tratamiento del lenguaje natural, incluso ofrecen la posibilidad de mejorar las búsquedas afinando los conceptos a buscar, pero continúan teniendo el resto de problemas, incluyendo el de la complejidad en su uso.

Finalmente, herramientas encaradas a sistemas basados en FAQs, como son Q-GO⁴ o las ofrecidas por Artificial Solutions⁵, presentan ventajas en cuando a la facilidad de uso, la concreción de las respuestas e incluso el cumplimiento de la LOPD y la imparcialidad esperada, en cambio se detecta un claro problema con respecto a su escalabilidad. A la hora de trabajar con un conjunto elevado de respuestas este sistemas son inviables debido al gran gasto en tiempo que hace falta por ajustarlos y para su mantenimiento, por la misma razón del coste operacional a la hora de añadir una respuesta, además del problema relacionado con la vigencia de la información, si se tiene en cuenta que se trabaja con grandes volúmenes de información.

2. Motor de búsqueda propuesto

Una vez un ciudadano realiza un pregunta, por ejemplo por el canal Web (véase Figura 2), iSAC realiza un tratamiento que presenta tres fases claramente diferenciadas. En la primera fase se hace un tratamiento del lenguaje natural, a continuación se calcula el grado de acierto de la pregunta con las respuestas posibles y finalmente se selecciona la respuesta o respuestas que se mostrarán al ciudadano.

3 <http://vivisimo.comhtmlvelocity>

4 <http://www.q-go.nl5300default.aspx>

5 <http://www.artificial-solutions.com>

Figura 2. Ejemplo de realización de pregunta por el canal Web



¿Cómo os podemos ayudar?

¿que hay que hacer para encender un fuego en un descampado?

Enviar

2.1. Tratamiento del lenguaje natural

En la fase de tratamiento del lenguaje natural se hace la eliminación de stopwords, una corrección ortográfica y el análisis morfológico. Hasta aquí, los pasos seguidos son básicamente los que se hacen en cualquier proceso de *text mining* o *information retrieval* [8] [9] [10]. Después se identifican patrones del lenguaje, se aplica una red semántica desarrollada especialmente para las administraciones públicas, se localizan expresiones con múltiples palabras y se identifican fechas de uso común. La red semántica representa todas las relaciones semánticas conocidas que hay entre las palabras, relacionando de este modo dos palabras aparentemente diferentes, aportando más calidad al análisis y tratamiento del lenguaje natural, esta red se basa en análisis previos del lenguaje y mejoras a lo largo del funcionamiento del sistema.

La parte referente a los *stopwords* se basa en la identificación y siguiente eliminación de todas aquellas palabras sin carga de información útil (p.ej: artículos, preposiciones,...). A continuación, usando herramientas de código abierto, como los diccionario de OpenOffice⁶ o la herramienta JMySpell⁷, se hace una corrección ortográfica del texto, analizándolo y proponiendo la corrección que más se adecue al lenguaje empleado por la ciudadanía. El siguiente paso es el análisis morfológico, extrayendo los lemas de las palabras de la pregunta mediante los diccionarios de la herramienta Freeling⁸ [11].

La parte más específica de nuestro motor empieza con el tratamiento de patrones del lenguaje que pueden emplear la ciudadanía en el momento de plantear una cuestión, entendiendo por patrón estructuras de palabras usadas de forma común en una pregunta, que nos aporta concreción y focaliza el objetivo que ésta tiene. Tras identificar todos los patrones existentes en la cuestión planteada se hace el tratamiento oportuno para cada caso.

A continuación se aplica la red semántica construida, localizando barbarismos (palabras de otras lenguas utilizadas comunmente), localismos (palabras propias de la región dónde se aplica el sistema), sinónimos, hipónimos, hiperónimos (parte - todo) o

⁶ <http://lingucomponent.openoffice.org/>

⁷ <http://jmyspell.javahispano.limpio/se/index.html>

⁸ <http://garraf.epsevg.upc.se/freeling/>

incorrecciones (palabras habitualmente mal escritas por los ciudadanos) y se relacionan estos con el significado raíz del diccionario, construyendo de este modo lo que nosotros hemos denominado bloque semántico, que representa un único valor semántico con diferentes representaciones del mismo: la palabra original, su lema o las relaciones encontradas a partir de la red semántica.

Finalmente se identifican expresiones de múltiples palabras (p.ej. “casa consistorial” que se utiliza para referirse al ayuntamiento), poniendo énfasis a las que son específicas del mundo administrativo y entre la ciudadanía no son tan utilizadas (p.ej. “unión consensual” es popularmente denominado “pareja de hecho”), y fechas de uso común (p.ej. ayer, este fin de semana o San Juan), se añaden estas a la búsqueda y de este modo se aporta una mayor cantidad de información útil al sistema.

2.2. Cálculo del grado de acierto

Antes de detallar la siguiente fase hace falta remarcar que, el conjunto de posibles respuestas de las cuales se dispone (no se ha de olvidar que se trata de una herramienta aplicada al servicio de atención ciudadana), se podría clasificar de forma clara en varios contextos diferentes, de la misma forma que se podrían clasificar las cuestiones planteadas por los ciudadanos con los mismos contextos. Por ejemplo, contextos comunes en servicios de atención ciudadana son el de trámites administrativos, el del directorio de entidades de la ciudad o el de actos de la agenda ciudadana.

De forma contraria a otros buscadores que acceden a la Web con la intención de extraer la información, el motor propuesto hace uso directo de las bases de datos del servicio de atención ciudadana, esto permite clasificar con mayor precisión toda aquella información a la cual se tiene acceso, construyendo un conjunto de información relacionada para cada respuesta basado en descriptores (palabras clave) de más calidad que la que permiten el resto de sistemas, pudiendo ajustar más la relevancia que tiene cada información dentro de la respuesta, dado que no aporta tanta información para una entidad el horario de trabajo como la descripción de la actividad que lleva a término, por ejemplo.

El grado de acierto entre la pregunta del ciudadano y una posible respuesta i ($tLevel_i$), se obtiene de la relación de los bloques semánticos construidos en la pregunta con los descriptores de la respuesta i , de si se ha accedido a la información relevante de la respuesta i , de la coincidencia entre el contexto de la pregunta y el de la respuesta i , y de la popularidad de la respuesta i [12] entre los ciudadanos.

$$\text{Eq. (1) } tLevel_i = \frac{\sum_{k=1}^n sTruth_{k,i} + \sum_{k=1}^m eTruth_{k,i}}{n + m} * q1 + iC_i * q2 + cC_i * q3 + rPopularity_i * q4$$

$sTruth_{k,i}$, obtenido de Eq.(2), es un valor que indica la coincidencia entre cada bloque semántico k de la pregunta y los descriptores de la respuesta i . n indica el número de bloques semánticos identificados en la pregunta. $eTruth_{k,i}$ indica la coincidencia entre cada expresión k de la pregunta y los descriptor de la respuesta i . m indica el número de expresiones identificadas en la pregunta. iC_i indica cuántos de los bloques semánticos y de

las expresiones estaban en posiciones relevantes dentro de la respuesta i . cC_i indica la coincidencia que tiene la búsqueda del ciudadano respecto el contexto de la respuesta i . $rPopularity_i$ es el factor que indica la popularidad de la respuesta i , se obtiene de Eq.(3). Los valores $q1, q2, q3, q4$ son pesos que dan más o menos importancia a los factores que intervienen en la fórmula. Estos pesos se ajustan durante la implantación del sistema. Es necesario remarcar que $q4$ es el peso que permite asegurar la imparcialidad del sistema cuando interese.

$$\text{Eq. (2)} \quad sTruth_{i,j} = (Entropy_i * q6 + q7) * Max(wW_{i,j}, lW_{i,j} * lP, sW_{i,j} * sP, bW_{i,j} * bP, pW_{i,j} * pP, hW_{i,j} * hP, mW_{i,j} * mP, iW_{i,j} * iP)$$

Donde $lEntropy_i$ que se calcula a partir de Eq.(4), indica la entropía (cantidad de información que puede aportar un elemento, y que se relaciona de forma inversa con las cantidad de repeticiones de este.) menor entre todos los elementos que conforman el bloque semántico. wW_e indica la coincidencia entre la palabra buscada con que se ha construido el bloque semántico i y los descriptores de la respuesta j . $lW_i, sW_i, bW_i, pW_i, hW_i, mW_i$ y iW_i son respectivamente los valores que indican el grado de coincidencia de los lemas, sinónimos, barbarismos, localismos, hipónimos, hiperónimos e incorrecciones pertenecientes al bloque semántico i con los descriptores de la respuesta j . lP, sP, bP, pP, hP, mP y iP son pesos que penalizan el grado de coincidencia dependiente de la relación semántica. $q6, q7$ son valores ajustables.

$$\text{Eq. (3)} \quad rPopularity_i = 1 - q5^{\left(\frac{clicks_i + throwings_i}{avgRconsults}\right)}$$

Dónde $clicks_i$ son el número de accesos voluntarios que hace un ciudadano a una respuesta i y $throwings_i$ el número de visualizaciones automáticas que se ha hecho de la respuesta i (ver 2.3). $avgRconsults$ indica la media de consultas de todas las respuestas, teniendo en cuenta para este cálculo $clicks$ y $throwings$. $q5$ es un parámetro que se ajusta igualmente durante la implementación del sistema.

$$\text{Eq. (4)} \quad entropy_x = \left| \log_2 \left(\frac{freq_i}{maxFreq} \right) / \log_2 \left(\frac{minFreq}{maxFreq} \right) \right|$$

$freq_e$ indica el número de repeticiones de la palabra i . $maxFreq$ indica la frecuencia de la palabra con mayor número de repeticiones. $minFreq$ indica la frecuencia de la palabra con menor número de repeticiones.

2.3. Selección de la respuesta

La última fase de la búsqueda hace referencia a la selección de la respuesta o respuestas que se mostrarán al ciudadano. Antes de nada se ordenarán las respuestas a partir del grado de acierto y se evaluará si la primera es lo suficiente precisa y, por lo tanto, se pueden obviar el resto, de lo contrario se hará un corte del número de respuestas.

Una vez se tienen ordenadas las respuestas a partir del grado de acierto, se hace una comparación entre la respuesta más relevante y el resto. Si la primera respuesta tiene una calidad mínima marcada y además supera con creces el resto, esta será lanzada de forma automática, de forma que el ciudadano sólo verá una respuesta.



Si no se ha lanzado ninguna respuesta automáticamente, se captura el grado de acierto de la respuesta más relevante y a partir de este se hace una discriminación del resto, obteniendo un conjunto de respuestas menor pero de mayor calidad (véase Figura 3).

Figura 3. Resultados de la pregunta de ejemplo realizada por el canal Web

Se han encontrado 2 resultados. Tiempo en responder: 0.726 segundos.

¿Es correcta la respuesta?

Sí No

Resultados: ¿que hay que hacer para encender un fuego en un descampado?		Grado de acierto
 COMUNICACIO PER FER FOC		↑ ↑ ↑ ↑ ↑
 PERMIS PER FER FOC EN PERIODES DE RISC		↑ ↑ ↑ ↑ ↑

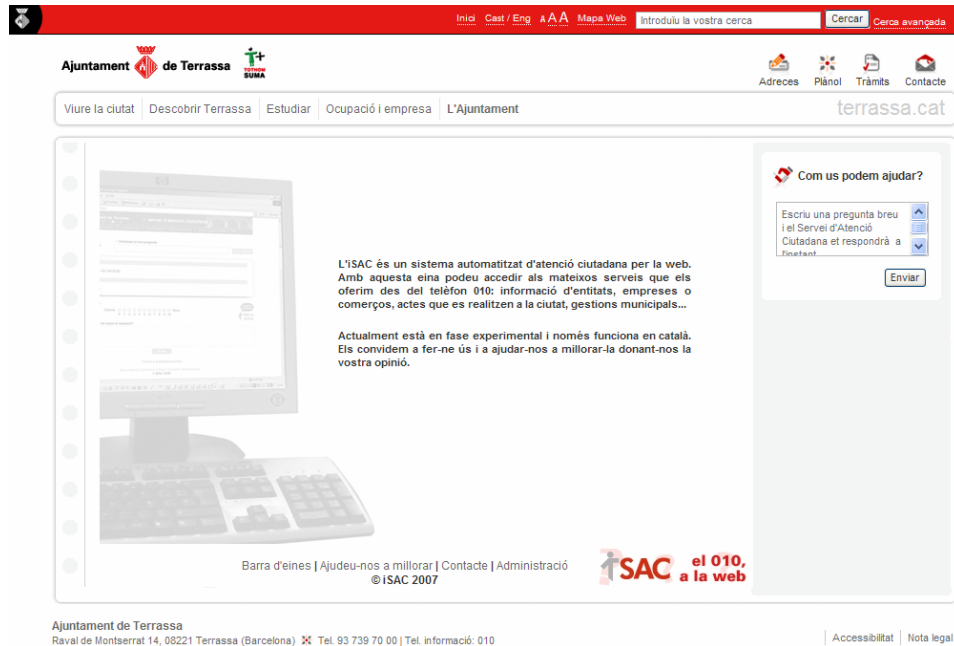
3. Experimentos y resultados

El objetivo de los experimentos que hemos llevado a término es demostrar que nuestro motor de búsqueda encuentra mejor las respuestas y es más concreto que el resto, y por lo tanto, es más adecuado para un servicio de atención ciudadana virtual. Las pruebas se han realizado entre las diversas herramientas que se pueden encontrar dentro de la Web del Ayuntamiento de Terrassa⁹ que es dónde se ha implantado iSAC¹⁰ (véase Figura 4).

⁹ <http://www.terrassa.cat>

¹⁰ <http://isac.terraza.org>

Figura 4 . iSAC en Terrassa (Barcelona)



Las comparaciones las hemos realizado contra el motor de búsqueda Google aplicado al portal del Ayuntamiento de Terrassa¹¹, el buscador de las administraciones desarrollado por la Administració Oberta de Catalunya¹², y un buscador propio desarrollado desde el ayuntamiento que se utilizaba antes de la implantación de Google¹³.

El experimento se ha llevado a cabo utilizando 1.200 preguntas reales realizadas en el servicio de atención ciudadana de Terrassa entre el 1 y el 10 de abril, del total de más de 62.000 preguntas realizadas de Enero a Octubre de 2007.

Este conjunto de preguntas se ha lanzado a las diferentes herramientas y se han comparado los resultados según los siguientes criterios: media de respuestas devueltas por el sistema (# respuestas), que nos indicará la concreción del mismo, y el porcentaje de acierto (% de acierto), donde consideramos acierto cuando la respuesta correcta se encuentra entre los 5 primeros resultados mostrados por la herramienta.

11

<http://www.google.com/custom?hl=se&domains=terrazza.limpio&q=&btnG=Buscar&site=terrazza.limpio>

12

<http://buscador.aocat.limpio/buscador.asp?Origen=20&tipo2=C&OrigenCerca=08279&organismo=08279>

13 <http://www.terrazza.cat/cercador>

Tabla 1 . Tabla de resultados:

	*iSAC	Google	Buscador *AOC	Buscador Propio
% de acierto	87%	58%	60%	38%
# resultados	23	547	455	18

En la tabla 1 se puede comprobar como en el caso del buscador propio se han obtenido pocos resultados con las buscas realizadas, pero por el contrario los aciertos han sido muy bajos. Esto es debido a que este buscador no hace el menor asomo de tratamiento de lenguaje natural y esto limita mucho el espacio de búsqueda de la herramienta.

En el caso de Google y del buscador AOC los resultados son parecidos, algo mejor en el caso del buscador AOC puesto que ésta es una herramienta desarrollada específicamente para su aplicación en administraciones públicas como ayuntamientos o consejos comarcales. En estos dos casos el porcentaje de aciertos sube considerablemente, pero debido a la filosofía con la que se desarrollan estas herramientas, que buscan la exhaustividad de resultados, el número de respuestas mostradas es muy elevado, en los dos casos alrededor de 500, y este hecho es contrario al funcionamiento que se pretende dar desde un servicio de atención ciudadana. Otro problema detectado es el de la imparcialidad, puesto que estos sistemas se basan en rankings por votos sin tener en cuenta esta problemática.

Finalmente, el motor de búsqueda propuesto encuentra la respuesta correcta en el 87% de los casos, este valor es un indicador muy buen tratamiento que se hace de la información. Por otra parte, el número medio de resultados que se muestran es de sólo 23. Aunque es muy menor que el del buscador AOC y que el de Google, consideramos que este valor todavía es elevado, puesto que la herramienta óptima sólo devolvería un único resultado y este sería el correcto, en la emulación más literal de una conversación telefónica de las que habitualmente se producen en los servicios de Atención Ciudadana - 010.

Es justo decir que los tiempos de respuesta de las diferentes herramientas son parecidas en todos los casos y estos valores son dentro del que los expertos en usabilidad han estimado como tiempo medio de espera de los usuarios [13].

4. Conclusiones y líneas futuras

Este artículo presenta la experiencia del proyecto iSAC, un servicio de atención virtual multicanal que incorpora un motor de búsqueda diseñado expresamente para las administraciones públicas y que ya está en plena explotación en la ciudad de Terrassa. Este sistema incorpora un motor de búsqueda que sigue un procedimiento de tres fases: tratamiento del lenguaje natural, clasificación de las posibles respuestas y selección de la respuesta/s correctas. Éste motor se ha comparado con otras herramientas ya existentes en la Web del Ayuntamiento de Terrassa, donde se pueden encontrar todas funcionando y, por lo tanto, permite una comparación válida. Los resultados observados son que, aun cuando otras herramientas consiguen un porcentaje de acierto elevado, presentan las carencias en cuando a concreción, LOPD o imparcialidad que se habían detectado al inicio del proyecto.

El motor propuesto en este documento presenta un alto acierto y una concreción suficiente para considerarlo una buena herramienta para la atención ciudadana.

Pensando en líneas futuras de trabajo relacionadas, es importante destacar que muchas veces los ciudadanos realizan preguntas a una administración pública sobre temas que no son competencia de ésta. Al no obtener una respuesta, el ciudadano queda descontento con el servicio recibido. Para hacer frente a este problema, se quiere desarrollar un sistema de comunicación entre los motores de búsqueda instalados en las diferentes administraciones (ayuntamientos, consejos comarcales, diputaciones, gobiernos autonómicos,...), accediendo a cualquier información que fuera de interés para el ciudadano, de forma transparente y sin importar de quien es competencia.

El camino también pasa por la gestión de contenidos basados en la filosofía WIKI[14], estando en mano de los propios usuarios el funcionamiento del sistema, con lo cual este se adaptaría más y mejor a las necesidades que presenta la ciudadanía.

Reconocimientos

Este trabajo se ha llevado a cabo gracias al apoyo del programa ICREA Júnior Empresa.

Referencias

- [1] *Atenció ciutadana, Dossier de premsa.* (2006)
http://www10.gencat.net/dursi/generados/catala/societat_informacio/recurs/doc/dossier012.pdf
- [2] *LEY ORGÁNICA 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal.* (1999)
<http://www.msc.es/organizacion/sns/planCalidadSNS/pdf/transparencia/LOPD19992.pdf>
- [3] M. Castells, Universitat Oberta de Catalunya; *Internet y la Sociedad Red*; Llició inaugural del programa de doctorat sobre la societat de la informació i el coneixement. (2001)
- [4] J. Grau, J. Guayar; *El negocio de buscar en Internet. Análisis de mercado de los buscadores 2003*; El profesional de la información, ISSN 1386-6710, vol. 13, Nº 4, pags. 292-300. (2004)
- [5] S. Brin, L. Page; *The Anatomy of a Large-Scale Hypertextual Web Search Engine*; WWW7 / Computer Networks 30(1-7), pags. 107-117 (1998)
- [6] S. Koshman, A. Spink, B. Cansen; *Web searching on the Vivísimo search engine*; Journal of the American Society for Information Science and Technology, Volume 57, pags. 1875 – 1887, ISSN:1532-2882. (2006)
- [7] Steinbach, M., Karypis, G., & Kumar, V; *A comparison of document clustering techniques*; KDD Workshop on Text Mining. (2000)
- [8] F. Verdejo, J. Gonzalo, D. Fernández, A. Peñas, F. López; *ITEM: un motor de búsqueda multilingüe basado en indexación semántica*; Proceedings JBIDI. (2000)
- [9] F. López-Ostenero, J. Gonzalo, F. Verdejo; *Búsqueda de información multilingüe: estado del arte*; Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial, vol. 22, pags. 11 – 35. (2003)
- [10] A. Broder; *A taxonomy of web search*; SIGIR Forum, Vol. 36, No. 2, pp. 3-10. (2002)
- [11] A. Gelbukh, G. Sidorov; *Analizador Morfológico Disponible: un recurso importante para PLN en español*; Taller de Herramientas y Recursos Lingüísticos para el Español y el Portugués. (2004)
- [12] S. Mizarro; *Quality control in scholarly publishing: A new proposal*; Journal of the American Society for Information Science and Technology. ISSN 1532-2882, vol. 54, pags. 989-1005. (2003)
- [13] Jakob Nielsen; *Response Times: The Three Important Limits*; ISSN 1548-5552. (1994)
- [14] A. Moreno, J. de la Rosa, C. Carrillo; *WIKIFAQ: Obtaining complete FAQs*; Frontiers in Artificial Intelligence and Applications—AI Research & Development. ISSN 0922-6389, vol. 146, pag. 283-290, IOS Press. (2006)