**Implementación de una aplicación web social para el etiquetado de fragmentos de videos de YouTube aplicando el concepto de ‘juegos con un propósito’**

Carla Salvatierra Arellano**( 1 )**, Félix Rivas Tomalá**(** **2 )** y Jorge Rodríguez Echeverría**(** **3 )**

Facultad de Ingeniería en Electricidad y Computación

Escuela Superior Politécnica del litoral (ESPOL)

Campus Gustavo Galindo, Km. 30.5 vía Perimetral

Apartado 09-01-5863. Guayaquil, Ecuador

csa89sm @gmail.com **( 1 )**, frivasto[@gmail.com](mailto:csa89sm@gmail.com) **(** **2 )**,  jirodrig@espol.edu.ec **(** **3 )**

**Resumen**

*El presente trabajo muestra el diseño e implementación de una aplicación Web en cuyo modelo se ha aplicado el concepto de “Juegos con un propósito” y utilizado “Input-Agreement” para generar descripciones de texto de manera automática a partir de material audiovisual. La principal diferencia con otros frameworks para marcar anotaciones de material multimedia es que el juego colecta etiquetas asociadas a marcas de tiempo con una fracción del costo que requeriría esta operación al contratar usuarios expertos. Como resultado de las pruebas, se observa que las marcas de anotaciones, evaluadas por un ser humano, resultan relevantes al contexto del video. Esto permite comprobar la efectividad del paradigma denominado “Human Computation”.*

*Una última importante aportación del trabajo expuesto se refiere al hecho de que el modelo desarrollado puede ser utilizado para la generación de etiquetas asociadas a videos en cualquier idioma. En particular, las pruebas generadas, produjeron etiquetas para videos en español.*

**Palabras claves**: *Juegos con propósito, etiquetas, videos, Human Computation*

**Abstract**

*This research presents the design and implementation of a Web application in which model it has been applied the concept of “Games with a purpose”, and “Input Agreement” has been used to generate text descriptions automatically from audiovisual material. The main difference with other multimedia tagging frameworks is that the game collects tags associated to timestamps with a fraction of the cost that this operation would cost if hired professionals would do it. As a result of the testing, it can be observed that the annotated tags, being evaluated for a human, are relevant to the video context. This could be used for proving the effectiveness of the paradigm called “Human Computation”.*

*A last important contribution of this work refers to the fact that the model developed can be used to generate tags in any language. In particular, the generated test, produced tags for videos in Spanish.*

**Key words**: *Games with a purpose, tagging, videos, Human Computation*

1. **Introducción**

El avance de la tecnología ha permitido, que hoy en día la solución a muchos problemas sea obtenida por un algoritmo corriendo en un computador. Sin embargo, existen aún problemas que sólo los humanos pueden resolver. Estos problemas se refieren especialmente a tareas que requieren de la creatividad y percepción humana para su solución, como la generación de descripciones de texto para diferentes marcas de tiempo en un video. Modelos de aplicaciones para combinar aportaciones de humanos y de algoritmos han sido propuestas en el área de Human Computation [1]. Otro concepto importante de notar es uno de los propuestos en mecanismos de Crowdsourcing, el cual afirma que la inteligencia colectiva de grupos de personas genera mejores resultados que los de un individuo experto [2]. Dentro de esta misma área se ha desarrollado el concepto de “Juegos con un propósito” o GWAP (Games with a purpose) por sus siglas en inglés. Los GWAP buscan aprovechar el gran número de personas que juegan en línea para resolver problemas donde cada participante, mientras se divierte, contribuye sin saberlo a realizar una parte del procesamiento de la solución.

Desde el primer GWAP propuesto por von Ahn denominado ESP [4], se han desarrollado varias aplicaciones como Peekaboom [5], Verbosity [6] y TagATune [7], que permiten tanto el etiquetado de imágenes, como de archivos de audio. Aun así, en la actualidad, no existe una alternativa para la generación automática de descripciones de videos, asociadas a intervalos de tiempo predefinidos. Aplicando estos conceptos y con el objetivo de contribuir al etiquetado de recursos educativos, esta investigación busca diseñar una aplicación Web que pueda utilizar el poder de atracción de los juegos en línea para obtener etiquetas referentes al contexto desarrollado en un determinado fragmento de vídeo.

1. **Tecnologías**

La idea de los GWAP está fuertemente ligada a problemas de Inteligencia Artificial que aún presentan dificultades para ser automatizados por el computador. En este capítulo se explican los fundamentos teóricos detrás de los conceptos de Human Computation y GWAP.

**2.1 Human Computation.**

Human Computation es un paradigma que utiliza el poder de procesamiento del cerebro humano, aprovechando sus destrezas y habilidades de manera original, para resolver problemas que las computadoras no han podido resolver [1]. Otra definición que se puede citar es la de Yang et al. [8] quienes definen a Human Computation como el proceso computacional que involucra a los humanos en ciertos pasos.

**2.2 Juegos con un propósito (Games with a purpose).**

Los GWAP son una clase de sistemas de Human Computation, capaces de obtener información útil a través de mecanismos que involucren a los juegos.

Hay tres plantillas básicas o mecanismos que describen la estructura de un GWAP [3]:

1. **Output-Agreement.**
2. **Input-Agreement.**
3. **Inversion-Problem.**

En Output-Agreement se escoge a 2 jugadores aleatoriamente entre los disponibles. Se les da la misma entrada y deben producir resultados basados en ella tratando de producir la misma respuesta de sus compañeros. Los jugadores no comparten sus respuestas, ni pueden comunicarse entre ellos. Deben coincidir en sus respuestas, no al mismo tiempo, pero en algún momento.

En Input-Agreement se escoge a 2 jugadores aleatoriamente entre los disponibles. Se les da una entrada conocida por el juego que puede ser “Igual o Diferente” y deben producir resultados que describan las entradas dadas, de tal manera que sus compañeros puedan adivinar si son las mismas o no. Los jugadores aquí, si intercambian sus respuestas.

En Inversion-Problem se escoge a 2 jugadores aleatoriamente entre los disponibles. A un jugador se le asigna el rol de “descriptor” y al otro como “adivinador”. Las respuestas del descriptor deben ayudar al compañero a determinar la respuesta original.

**2.3 Node.js y Socket.io (WebSockets)**

Éste estándar simplifica mucho la complejidad de la comunicación Web bidireccional y el manejo de las conexiones. El protocolo para el lado servidor está siendo normalizado por el Grupo Especial sobre Ingeniería de Internet o IETF, por sus siglas en inglés, y el API de Websockets para el lado cliente está siendo estandarizado por la W3C.

Node.js es un entorno de ejecución y una librería, que usa un modelo basado en eventos, entradas y salidas no bloqueantes y por lo tanto asíncrona, ideal para uso intensivo de datos en aplicaciones de tiempo real que se ejecutan a través de dispositivos distribuidos.

1. **Calidad de las etiquetas**

Los diseñadores de juegos de Human Computation se enfrentan con el desafío de construir un sistema que cumpla simultáneamente dos objetivos: satisfacer a los jugadores y llevar a cabo un procesamiento eficiente y preciso. Por un lado, se puede diseñar un juego que atraiga a muchos jugadores, pero que no recoja datos útiles, y por otro lado, si el juego está diseñado para recopilar los datos más limpios posibles, sin prestar atención a cuán divertida es la tarea, entonces ningún jugador estaría interesado en participar.

Lo mejor es diseñar mecanismos en que los jugadores sean naturalmente motivados. Otras formas de obtener buenos resultados son las observadas en juegos como el ESP y KissKissBan. En el ESP, se aplica por ejemplo, las denominadas “taboo-words” o palabras bloqueadas, como una forma de obligar al jugador a ser más creativo en el ingreso de etiquetas, impidiendo así, el ingreso de aquellas muy generales [4].

Algo parecido se implementa en KissKissBan [11], con la variante de que ya no son sólo parejas de jugadores, sino que se introduce un nuevo rol, un tercer jugador que actúa como el moderador de las etiquetas, impidiendo a los otros dos el ingresar aquellas que él indique. Esto además hace que el juego sea competitivo y colaborativo a la vez, y que el control de palabras bloqueadas dependa del criterio de otro humano.

Estas dos soluciones son una forma de decirle explícitamente al usuario, que esas son las palabras que no debe ingresar. Tras observar estas ventajas del mecanismo “Output-Agreement”, cabe indicar que los mecanismos a pesar de ser generalizaciones, no son aplicables a todos los problemas. Por ejemplo, el mecanismo “Output-Agreement” se basa en el hecho de que es fácil conseguir coincidencias de etiquetas para una imagen [9], pero no así para videos, música o sonidos, como es el caso en el que mejor se ajustan los juegos basados en “Input-Agreement”.

**3.1 Estrategias para evitar acuerdos entre jugadores**

Herdagdelen y Baroni [10] indican que una forma de evitar que los jugadores hagan trampas es introduciendo hechos cuyas respuestas ya sean conocidas por el sistema, con el fin de determinar si un jugador está verdaderamente participando en el juego. Sin embargo, esto requiere una gran colección de hechos comprobados que no son fáciles de obtener. Por ello lo mejor es diseñar mecanismos en que los jugadores sean naturalmente motivados y jueguen sin malas intenciones [9].

Von Ahn y Dabbish [4] presentan varios mecanismos para contrarrestar las trampas en el juego ESP y son los siguientes:

● El servidor del juego inicia una nueva partida cada 30 segundos: cuando un nuevo jugador se autentica, debe esperar hasta el siguiente límite de 30 segundos para emparejarlo con otro jugador y empezar su juego. Esto se hace para asegurar que los jugadores se emparejen al azar y no logren hacer trampas autenticándose al mismo tiempo que sus compañeros.

● Como los jugadores forman parejas aleatorias, la probabilidad de que estos se pongan de acuerdo en un estrategia autenticándose al mismo tiempo es baja, porque no tendrán información sobre quién es su pareja.

● Midiendo el tiempo promedio en que los jugadores ingresaron sus etiquetas puede reflejar si se ha intentado hacer trampas estratégicamente entre varios jugadores, pues una marcada disminución del tiempo promedio debe indicar un acuerdo masivo en una estrategia.

**3.2 Incentivos**

Otro factor importante en los sistemas de Human Computation son los incentivos, los cuales afectan el comportamiento de los participantes incluyendo: si realizan la tarea en su totalidad, o nivel de participación, la tarea que escogen y lo bien que la realizan [9].

Los incentivos pueden ser de 2 tipos, según la forma de recompensar al participante:

Aquellos provocados por Motivaciones intrínsecas: sea por altruismo (por el beneficio de otros), el deseo de aprender o por competencia.

Motivaciones extrínsecas: como las recompensas monetarias (virtuales o reales) o por puntos y otras mecánicas de juegos.

1. **Análisis y Diseño de la aplicación**

El juego se diseñará de tal manera que varios jugadores puedan producir metadatos de forma conjunta. Dado que los usuarios accederán a la aplicación desde diversas plataformas, y que el contenido a etiquetar estará disponible “en línea”, es importante que el juego sea diseñado como una aplicación Web. La solución también requiere utilizar contribuciones de grandes cantidades de usuarios, por lo que se aprovecharán los conceptos asociados a GWAP, específicamente el mecanismo “Input-Agreement” por ser éste ideal para obtener información sobre recursos de alto contenido descriptivo como videos, sonidos y música.

La aplicación deberá conseguir:

● La autenticación de un usuario a través de Facebook.

● Crear una nueva partida.

● Asignar automáticamente compañeros de juego, es decir, dos jugadores por sesión. (Donde, el máximo está determinado por el número de sesiones soportadas en el servidor).

● Asignar un conjunto de fragmentos de video para las rondas de la partida a iniciar.

● Reproducir un fragmento de vídeo de YouTube para ambos jugadores de manera sincrónica. Dicho video puede ser aleatoriamente igual o diferente entre compañeros, pero debe pertenecer a una misma categoría.

● Mostrar una sección para que los jugadores ingresen palabras o frases describiendo el fragmento de vídeo actual, siempre que el jugador cuente con tiempo disponible en la partida.

● Visualizar las palabras o frases descritas por su compañero de juego en tiempo real.

● Mostrar un control para permitir al jugador decidir si el vídeo es igual o diferente al de su compañero.

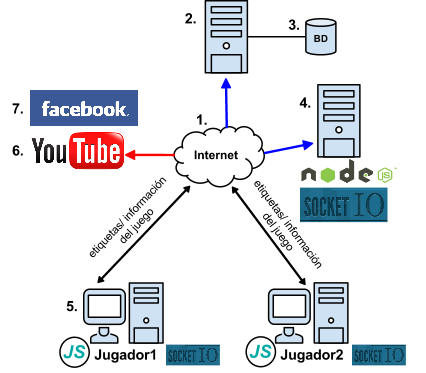
● Acumular automáticamente puntos cada vez que ambos compañeros están de acuerdo en la decisión sobre el video.

● Mostrar los resultados del juego, puntajes y resumen del mismo al finalizar una partida.

● Visualizar las bonificaciones de puntos al llegar a un determinado nivel en los puntajes acumulados.

● Terminar la sesión de juego en el momento en el que los jugadores lo deseen.

1. **Diseño General**



***Figura 1: Diseño lógico de ‘CazaVideos’.***

La aplicación a desarrollar será implementada utilizando un lenguaje de programación Web que permita generar páginas de forma dinámica. El control de las acciones del juego será implementado en diversos módulos según la funcionalidad implementada. A continuación, se enumeran dichos módulos y se detalla el objetivo de cada uno de ellos:

● Elaboración de la Mesa.

● Control de la Mesa.

● Sincronización de interfaces.

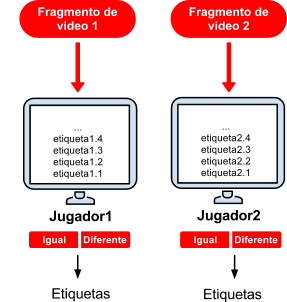
Elaboración de la Mesa: Módulo en el que se incluye la lógica a implementar para iniciar una sesión de juego. Esto incluye tanto la autenticación como las rutinas para asociar en parejas los usuarios disponibles.

Control de la Mesa: Módulo en el que se implementarán las rutinas que controlan los eventos que se disparan durante el desarrollo del juego. Entre las rutinas a implementar están las validaciones para decidir si el juego continúa o no, el manejo de tiempo, la acumulación de puntos y los cambios de estado.

Sincronización de interfaces: En este módulo se implementará el backend que permite la sincronización de mensajes y decisiones entre las parejas de jugadores. Para la implementación se utilizará el método de comunicación conocido como “WebSockets” [12], determinado en el siguiente análisis.

En la Figura 1 se puede observar el flujo de información requerido para el desarrollo del juego. Los clientes entre sí y con el servidor intercambian: mensaje o etiqueta, tiempo de juego, vídeo y decisión. Otro factor importante a notar es que la aplicación no almacena los vídeos, estos son tomados directamente de los servidores de YouTube y presentados en la interfaz del juego.

En la Figura 2, se representa la aplicación del concepto “Input Agreement” como una solución al sistema, en el que las entradas representan a los fragmentos de vídeo. El proceso es la acción de los jugadores de producir etiquetas asignadas a sus respectivos fragmentos de vídeo y las salidas son las etiquetas que se generan a medida que transcurre el tiempo del juego.



***Figura 2: Mecanismo de “Input-Agreement” aplicado en ‘CazaVideos’.***

**5.1 Filtrado de la información**

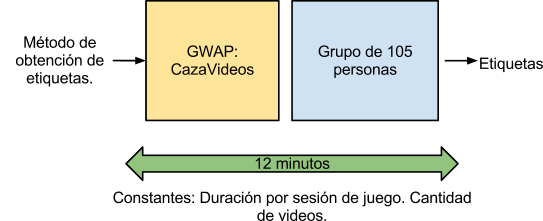
El ingreso de etiquetas no se restringe a un vocabulario predefinido. Se permitirá el ingreso de todas las palabras escogidas por el jugador y se las filtrará al final realizando un post-procesamiento para obtener las etiquetas oficiales. Cabe indicar que son muy pocas las etiquetas que se usan para hacer trampa, pues estas normalmente resultan de ponerse de acuerdo con el compañero usando el intercambio de mensajes como medio de comunicación [9]. El hecho de que las parejas de jugadores sean escogidas aleatoriamente y que cada fragmento de vídeo sea reproducido por únicamente 30 segundos, son también una forma de garantizar que los usuarios no podrán comunicarse entre sí. Al final se presentará a los usuarios un resumen del puntaje de su participación durante la sesión de juego, con la finalidad de que puedan aprender de sus errores.

Se implementará una rutina que agrupe las etiquetas en función del fragmento de vídeo, luego a cada grupo se aplican algoritmos para excluir anotaciones que no sean relevantes. Para efectos del alcance de este trabajo, los datos que se filtrarán serán: caracteres especiales y palabras de apoyo a la oración, conocidas como Stop Words o Palabras Vacías, como por ejemplo los artículos, preposiciones, etc. Adicionalmente se utilizará el proceso de stemming, el cual permite detectar etiquetas cuya intención sea describir la misma palabra, como por ejemplo las conjugaciones de verbos.

Cabe decir que la gran cantidad de datos redundantes facilita encontrar anotaciones válidas entre aquellas que tienen frecuencia más alta. Las etiquetas finales serán presentadas como nubes de tags donde se verán resaltadas las palabras con mayor frecuencia.

**5.2 Diseño de plan de pruebas**

Esta investigación se llevará a cabo usando un diseño cuasi-experimental ya que la muestra fue escogida de forma no aleatoria, conformada por estudiantes de ingeniería en su mayor parte. En la Figura 3 se puede observar el tiempo que cada usuario tendrá para generar las etiquetas, el cual corresponde a dos sesiones de juego, o 12 minutos de duración.



***Figura 3: - Diseño cuasi-experimental***

Se prevé que los usuarios trabajen 16 horas, el mismo número de horas requerido que una persona se demoraría en etiquetar únicamente los fragmentos educativos, a las que, para prevenir errores con un 76% de confianza aproximando a 21 horas de trabajo, se les agrega 5 horas adicionales. Por lo tanto se escogerán 105 personas, cada una jugará las dos sesiones de juego establecidas, para probar la capacidad de generación masiva de etiquetas

Los fragmentos de vídeo pueden ser de tres tipos: “Educativos”, “Blooper” y “Popfail”. En la categoría Educativos, los vídeos que se extraerán serán obtenidos del canal de YouTube de ESPOL, los cuales presentan sesiones de clase de las áreas: física, matemáticas o química. Los de tipo Blooper, presentan accidentes casuales y los de tipo Popfail, presentan situaciones inesperadas durante interpretaciones de artistas populares. Las anotaciones generadas en el experimento serán agrupadas según su fragmento de vídeo y su frecuencia, y de estas se medirá visualmente la calidad basado en su relevancia con el tema y su generalidad.

1. **Pruebas y Resultados**

Se realizaron pruebas a los datos recogidos durante el tiempo en que estuvo la aplicación en producción recaudando un total de 5,529 etiquetas en un total de 210 sesiones de juego.

Las 105 personas escogidas fueron expuestas al uso de la aplicación. La edad media de esta muestra es 23 años, con nivel de educación superior, 67% de sexo masculino, 33% femenino y estado civil soltero en su totalidad.

***Pruebas de funcionalidad.***

Todas las pruebas propuestas en el capítulo 3.2.4 fueron realizadas en ordenadores con procesadores intel Core 2 duo, 2GB de memoria RAM, Sistema Operativo Windows 7, navegador Google Chrome y velocidad de Internet de 512 Kbps. A continuación en la Figura 4 se muestra la pantalla en la que dentro de los 30 segundos establecidos se realiza una anotación exitosa, las etiquetas fueron ingresadas correctamente

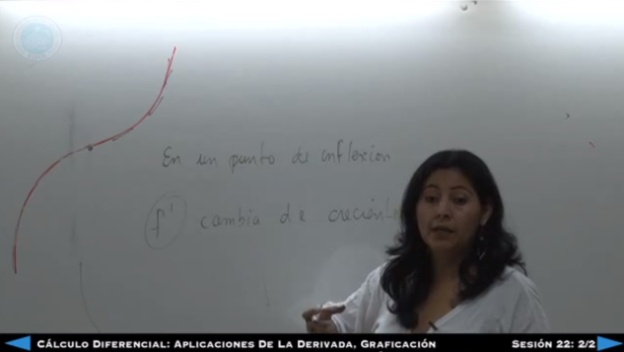


***Figura 4: Obtención de Anotaciones durante el Juego.***

Luego de estas pruebas, se empezó a tomar en cuenta dos requisitos, no contemplados en el diseño, que son: asegurarse de que los dispositivos usados para probar mantengan usables los puertos destinados para la aplicación y el Sincronizador, y que se cumpla con el ancho de banda mínimo requerido para un funcionamiento aceptable, en este caso, 512Kbps.

***Calidad de la información recolectada.***

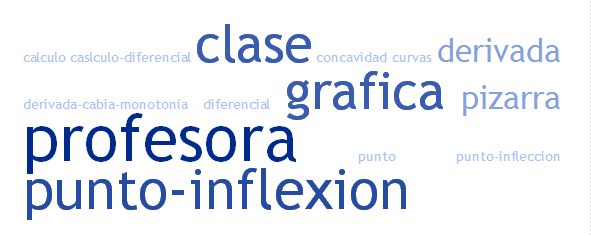
Para la medición de la calidad de la información, de un total de 30 vídeos de la categoría “Educación”, y como se explicó en la sección 5.1, las etiquetas que se almacenaron corresponden a fragmentos de vídeo de 30 segundos. Se tomaron en cuenta 36 fragmentos de dichos vídeos y se generó la nube de tags, un ejemplo de esto se muestra en las imágenes a continuación, donde la figura 5 muestra una clase de cálculo diferencial en el fragmento de tiempo entre los minutos 00:11:50 y 00:12:20. El tema principal tratado en este caso es “punto de inflexión”. Luego, en la figura 6 se puede observar la nube de tags generada como producto de las sesiones de juego invertidas específicamente en este ejemplo, y por último, en la 7 se muestran estas mismas etiquetas luego del proceso de filtrado descrito en la sección 5.1.



***Figura 5: Fragmento de video entre 00:11:50 y 00:12:20.***



***Figura 6: Etiquetas antes del filtrado.***



***Figura 7: Etiquetas después del filtrado.***

De la figura 7 también se puede visualizar que el par de etiquetas más importantes es “profesora” y “punto-inflexión”. El número de jugadores que tuvieron que participar hasta llegar a obtener etiquetas válidas es de 5 en promedio

***Análisis de los resultados***

Al observar las nubes de etiquetas de la figura 7, la etiqueta de mayor frecuencia es muy general, por tanto no agrega valor al etiquetado del fragmento. Por otro lado, omitiendo esta, se puede observar que efectivamente se logró etiquetas válidas, aportando realmente una descripción al tema que se está desarrollando, como por ejemplo: “punto-inflexión”.

El número de usuarios necesarios resultó ser impar, debido a que no necesariamente los usuarios etiquetan en par el mismo vídeo. Teniendo en cuenta que el promedio de usuarios es de 5 y que los usuarios podrían tardar hasta 30 segundos en realizar las anotaciones, podemos decir que un usuario se tardaría alrededor de 2 minutos y 50 segundos en etiquetar un fragmento de vídeo; mientras un humano, para etiquetar ese mismo fragmento, debería ver todo el contenido.

Las estadísticas del Juego se muestran en la Tabla I.

|  |  |
| --- | --- |
| Total etiquetas recolectadas: | 5529 |
| Total etiquetas únicas | 1514 |
| Sesiones de juego | 210 |
| Total de usuarios | 105 |

***Tabla I: Estadísticas del Juego.***

En las siguientes tablas se muestra la cantidad de etiquetas obtenidas luego del procesamiento. En la Tabla II, aquellas que hemos denominado bloqueadas por no aportar con información específica del contexto del fragmento de video, y en la Tabla III, la cantidad de etiquetas válidas, pertenecientes al intervalo 12, ver nube de etiquetas de la Figura 7.

|  |  |
| --- | --- |
| Cantidad de etiquetas | Usuarios de Facebook |
| 4 | Usuario1 |
| 1 | Usuario2 |
| 2 | Usuario3 |
| 2 | Usuario4 |

***Tabla II: Cantidad de etiquetas bloqueadas del intervalo 12.***

|  |  |
| --- | --- |
| Cantidad de etiquetas | Usuarios de Facebook |
| 4 | Usuario1 |
| 3 | Usuario3 |
| 1 | Usuario5 |

***Tabla III: Cantidad de etiquetas válidas por usuario del intervalo 12.***

1. **Conclusiones**

* Los GWAPs presentan un mecanismo eficaz para la obtención de etiquetas relevantes con cero costo en lo que se refiere a las personas que participan en el proceso y un costo mínimo de infraestructura para mantener la aplicación Web online.
* Es posible que una aplicación consiga, con relativamente pocos usuarios, mejorar en 10500% el tiempo en que se realiza la generación de etiquetas para material audiovisual educativo comparado con la contratación de un experto.
* Debido a que en el mecanismo aplicado, “Input-Agreement”, se permite el ingreso de todo tipo de etiquetas, para no ser tan restrictivos y más bien velar por la motivación del jugador, es necesario realizar el procesamiento offline, para el filtrado de las etiquetas ingresadas.

1. **Recomendaciones**

* Para futuras investigaciones en esta misma línea, se recomienda obtener videos de canales públicos o implementar un canal de YouTube propio del juego, así se podría evitar errores de tipo “100” de YouTube.
* En cuanto al GWAP, como próximos pasos, se sugiere experimentar con una mecánica de juego entre parejas no colaborativo, es decir, competitiva, ya que varios de los usuarios manifestaron que preferían recibir puntaje por “ganarle” a su pareja. Otra opción sería un híbrido colaborativo-competitivo, como lo implementa el juego “KissKissBan”.
* Para la implementación, es aconsejable minimizar el paso de mensajes, una forma de hacer esto podría ser evitando el uso de otro framework como cliente, sino que se maneje la lógica directamente en Node.js; En ese caso se puede usar un framework para Node.js y la base de datos para no albergar demasiadas variables en memoria. Existen bases de datos no relacionales para Node.js como Redis, CouchDB, MongoDB que aportarían un acceso rápido a datos en el juego.
* En cuanto a la generación de etiqueta final, se deja un camino trazado que, se recomienda, empezar mejorando el filtrado de la información, introduciendo técnicas de completado de raíces de palabras, como por ejemplo, usando Stemming Completion.

1. **Referencias**

[1] Von Ahn, L. Human computation. Tesis Doctoral. UMI Order Number: AAI3205378, CMU, 2005

[2] Surowiecki, James. The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations. Doubleday, 2004.

[3] Von Ahn, L. y Dabbish, L. General Techniques for Designing Games with a Purpose. Communications of the ACM, Agosto 2008. Páginas 58-67.

[4] Von Ahn, L. y Dabbish, L. Labeling images with a computer game. In Proc. SIGCHI Conf. on Human Factors in Computing Systems (Vienna, April 24-29). ACM, New York, 319-326, 2004.

[5] Von Ahn, L., Liu, R., and Blum, M. Peekaboom: A Game for locating objects in images. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Montreal, Apr. 22–27). ACM Press, New York, 2006, 55–64.

[6] Von Ahn, L., Kedia, M., and Blum, M. Verbosity: A game for collecting common-sense knowledge. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Montreal, Apr. 22–27). ACM Press, 2007, 75–78.

[7] Law, E., von Ahn L., Dannenberg, R., and Crawford M. TagATune: a game for music and sound annotation. In Proc. 8th Intl. Conf. on Music Information Retrieval (Vienna, September 23-27). 361-364, 2007.

[8] Yang Yang, Zhu Bin B., Guo Rui, Yang Linjun, Li Shipeng, and Yu Nenghai. 2008. A comprehensive human computation framework: with application to image labeling. In Proceedings of the 16th ACM international conference on Multimedia (MM '08). ACM, New York, NY, USA, 479-488.

[9] Law, E. y von Ahn, L. Human Computation. Morgan & Claypool Synthesis Lectures on Artificial Intelligence and Machine Learning, 2011.

[10] Herdagdelen A. y Baroni M. The concept game: Better commonsense knowledge extraction by combining text mining and a game with a purpose. In AAAI Fall Symposium on Commonsense Knowledge, 2010.

[11] Ho, C.-ju, Chang, T.-hsuan, Lee, J.-chuan, Hsu, J. Y.-jen, & Chen, K.-ta. (2009). KissKissBan: A Competitive Human Computation Game for Image Annotation Categories and Subject Descriptors. Memory, 1-4. ACM Press. Recuperado de http://portal.acm.org/citation.cfm?id=1600153

[12] Ubl, M. y Kitamura, E. (2010, Octubre 20) Introducing Websockets: Bringing Sockets to the Web. Recuperado de: http://www.html5rocks.com/en/tutorials/websockets/basics/