

Alumno: Nelzon Jorge Apaza Apaza.

1. Dataset

En el contenido del dataset se puede apreciar que se tiene la presencia de 1445 noticias (archivos xml), blogs, imágenes, databases (archivos csv, xls) y documentos PDF y DOC.

Noticias

[illegible]

```

final version" &f' encoding="UTF-8")
doc = etree.parse('2006-07-12 2003/notes
content)

MAMMISTAN - the prairie dog, African hedgehog and the Chinchilla, are among
No one should get a fat pet simply because it is "cute," Parmentier said.
"They might acquire a tiger cub, not realizing that its weight over 400
pounds.
The chinchilla, one of the "fat" pets of the moment, will select a diet far
Diet-related ailments are common among fat pets because nature don't know the
</content>
</doc>
```

[illegible]

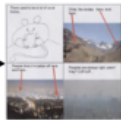
Cada archivo XML tiene dos elementos principales:

- **<date>**: una fecha y hora (ej. Tue May 20 06:07:21 2003)
- **<content>**: texto periodístico, informativo o narrativo

Propuesta de etiquetas (temas y categorías):

Unset

```
{
  "file": "doc123.xml",
  "date": "2003-05-20",
  "labels": [
    "Mascotas",
    "Mascotas exóticas",
    "Salud animal",
    "Artículo informativo",
    "Veterinario",
    "Oregon State University",
    "Advertencia"
  ],
  "animals": [
    "Perro de la pradera",
    "Erizo africano",
    "Chinchilla",
    "Tigre",
    "Pitón birmana",
    "Petauro del azúcar"
  ],
  "topics": [
    "Dietas inadecuadas",
    "Riesgos del animal",
    "Popularidad de mascotas"
```

[illegible]

- Se tienen imágenes en los mismos blogs y pocas en una carpeta "images"



The screenshot displays the Microsoft Dynamics CRM 4.0 interface. At the top, there is a navigation bar with tabs for 'Home', 'Sales', 'Marketing', 'Service', and 'Administration'. Below the navigation bar is a ribbon with various action buttons. The main area shows a list of 'Sales Opportunities' with columns for 'Name', 'Status', 'Owner', 'Created On', 'Modified On', 'Due Date', and 'Amount'. The list contains several entries, including 'Sales Opportunity 1', 'Sales Opportunity 2', and 'Sales Opportunity 3'. The 'Sales Opportunity 1' entry is highlighted, showing details such as 'Status: Open', 'Owner: John Doe', 'Created On: 10/10/2010 10:10:10 AM', 'Modified On: 10/10/2010 10:10:10 AM', 'Due Date: 10/10/2010', and 'Amount: 10000.00'.

[illegible]

Cada fila representa un **registro de permiso de importación de animales**, que describe un evento de importación específico, relacionado con: una especie animal, un número de individuos, el país de origen y destino, las personas o entidades involucradas, fechas relevantes (emisión y vencimiento del permiso), y detalles administrativos.

| Nombre (Inglés) | Traducción (Español) | Tipo de dato | Tipo estadístico | Descripción |
|----------------------|----------------------------|--------------|----------------------------------|---|
| Quantity | Cantidad de animales | Número | Cuantitativa discreta | Número total de animales importados. |
| Purpose | Propósito | Texto | Cualitativa nominal (codificada) | Código que representa el propósito de la importación. Ej. "C" puede significar "comercial". |
| Document_Title | Título del documento | Texto | Cualitativa nominal | Tipo de permiso, como "Import Permit". |
| Special_Conditions | Condiciones especiales | Texto | Cualitativa nominal | Información especial o restricciones (puede estar en blanco o "NA"). |
| Description | Descripción | Texto | Cualitativa nominal | Descripción general del envío (puede estar vacía). |
| Source_Code | Código fuente | Texto | Cualitativa nominal | Código de la agencia que origina el certificado. |
| Certificate_Number | Número de certificado | Texto | Cualitativa nominal | Número identificador único del certificado. |
| Valid_Until | Válido hasta | Fecha | Cualitativa ordinal (fecha) | Fecha de vencimiento del permiso. |
| Consignee_Country | País del consignatario | Texto | Cualitativa nominal | País que recibe el envío (consignatario). |
| Management_Authority | Autoridad de gestión | Texto | Cualitativa nominal | Entidad que emite el permiso. |
| Common_Name | Nombre común de la especie | Texto | Cualitativa nominal | Nombre común del animal. |

| | | | | |
|-------------------|--|-------|-----------------------------|--|
| Permittee_Country | País del solicitante | Texto | Cualitativa nominal | País del que solicita el permiso. |
| Consignee_Address | Dirección del consignatario | Texto | Cualitativa nominal | Dirección del receptor del envío. |
| Origin | Origen del envío | Texto | Cualitativa nominal | País o región de donde proviene el envío. |
| Permittee_Name | Nombre del solicitante | Texto | Cualitativa nominal | Persona o entidad que solicita el permiso. |
| Scientific_Name | Nombre científico de la especie | Texto | Cualitativa nominal | Nombre binomial de la especie (ej. Xerus rutilus). |
| Permittee_Address | Dirección del solicitante | Texto | Cualitativa nominal | Dirección del solicitante. |
| Consignee_Name | Nombre del consignatario | Texto | Cualitativa nominal | Persona o entidad que recibe el envío. |
| Issue_Date | Fecha de emisión del permiso | Fecha | Cualitativa ordinal (fecha) | Fecha en la que se emitió el permiso. |
| Security_Stamp | Sello de seguridad | Texto | Cualitativa nominal | Código de seguridad asociado al permiso. |
| USER WARNING | Advertencia (texto fijo: Synthetic Data) | Texto | Cualitativa nominal | Indica que los datos son sintéticos o simulados. |

Otros documentos

| Nombre | Fecha de modificación | Tipo | Tamaño |
|----------------------------|------------------------|----------------------|----------|
| MSDS | 23/02/2007 12:33 p. m. | Carpetas de archivos | |
| DEA Files Updatev2.doc | 27/02/2007 01:50 p. m. | Documento de Mi... | 1,847 KB |
| SFPixSum.pdf | 16/02/2007 01:38 p. m. | Vivado PDF Docu... | 70 KB |
| Trade.doc | 16/02/2007 01:38 p. m. | Documento de Mi... | 25 KB |
| Transport of Live Fish.doc | 16/02/2007 01:38 p. m. | Documento de Mi... | 122 KB |

LIQUID COCAINE INSIDE THE LININGS OF PLASTIC MUGS IN MIAMI, FLORIDA

The DEA Southeast Laboratory (Miami, Florida) recently received six colored acrylic/plastic mugs containing liquid inside the linings, suspected liquid cocaine (see Photo 1). The mugs (three gray/black, two clear, and one red/pink) were contained in an express mail package, and were interrogated by DEA Miami at the Miami International Airport Foreign Mail Facility. Analysis of the liquid (total net volume 1331 milliliters) by GC, GC/MS, IR, color testing, and urine testing confirmed 53 percent cocaine hydrochloride (510 milligrams/mliliters). This is the first encounter with this type of smuggling technique by the Southeast Laboratory.



CHRONIC TOXICITY SUMMARY
METHYL BROMIDE
(bromomethane; monobromomethane)
CAS Registry Number: 74-83-9

- I. Chronic Toxicity Summary**
- Inhalation reference exposure level* 5 µg/m³ (1 ppb)
Critical effect(s) Histological lesions of the olfactory epithelium of the nasal cavity in rats
Hazard index target(s) Respiratory system; nervous system; development
- II. Physical and Chemical Properties (HSDB, 1994)**
- | | |
|--------------------------|---|
| <i>Description</i> | Colorless gas |
| <i>Molecular formula</i> | CH ₃ Br |
| <i>Molecular weight</i> | 94.95 g/mol |
| <i>Density</i> | 3.89 g/L @ 25°C |
| <i>Boiling point</i> | 3.6°C |
| <i>Vapor pressure</i> | 1420 torr @ 20°C |
| <i>Solubility</i> | Soluble in ethanol, benzene, carbon disulfide, and 1,75% (w/w) in water |
| <i>Odor threshold</i> | 20.6 ppm |
| <i>Odor description</i> | Sweetish odor |
| <i>Miscibility</i> | Methanol, bromide, 5-methylcytosine |
| <i>Conversion factor</i> | 1 ppm = 3.89 µg/m ³ @ 25°C |

- III. Major Uses and Sources**
- Methyl bromide (MeBr) was used historically as an industrial fire extinguishing agent and was introduced in the U.S. from France in the 1920s. Current uses of MeBr include the fumigation

| Campo | Descripción | Ejemplo | Tipo estadístico |
|--------------------|---|---|-----------------------|
| Case_Title | Título o encabezado del informe | "HEROIN IN BOOK BINDINGS IN FORT LAUDERDALE, FLORIDA" | Cualitativa nominal |
| Date_Received | Fecha en que el laboratorio recibe/ingresa la muestra | 2003-05-20 | Fecha (ordinal) |
| Location | Ciudad / Estado / País donde ocurrió la incautación | Miami, Florida, USA | Cualitativa nominal |
| Laboratory | Nombre del laboratorio o agencia forense | DEA Southeast Laboratory | Cualitativa nominal |
| Agency_Seizing | Entidad que incautó el material | Broward Sheriff's Office | Cualitativa nominal |
| Submission_Type | Tipo de muestra (e.g., "Book binding", "Plastic mug", etc.) | Hard-cover books | Cualitativa nominal |
| Contraband | Sustancia(s) detectada(s) | ["Heroin HCl"] | Cualitativa nominal |
| Quantity_Total | Cantidad total incautada (masa o volumen) | 99.0 g | Cuantitativa continua |
| Purity_Percent | % de sustancia activa aproximada | 86% | Cuantitativa continua |
| Concealment_Method | Manera de ocultar la sustancia | Hollowed-out book binding | Cualitativa nominal |
| Analysis_Methods | Técnicas analíticas usadas | ["GC/MS", "FT-IR"] | Cualitativa nominal |
| Novelty | ¿Primera vez que se ve esta técnica? (Sí/No) | Sí | Cualitativa nominal |
| Photos_Referenced | Número de figuras mencionadas (Photo 1, Photo 2, ...) | 2 | Cuantitativa discreta |

| | | | |
|--------------|---------------------------------------|-------------|---------------------|
| Editor_Notes | Observaciones del editor (si las hay) | Texto libre | Cualitativa nominal |
|--------------|---------------------------------------|-------------|---------------------|

- Ejemplo de etiquetas temáticas y taxonomía:

Unset

Tipo de droga

Heroin
Cocaine
Salvia Divinorum
Psilocybin
Fentanyl
MDMA/PCP

Vehículo de contrabando

Book binding
Plastic mug
Lotion bottle
Chocolate candy
Baseball cap parts
Picture frame
Baby formula can
Wooden tray table

Técnica de análisis principal

GC/MS
GC-FID
FT-IR
Microscopic examination
Color testing
TLC

Riesgos / Observaciones clave

High purity
First encounter
Enhanced potency
New smuggling method
Health hazard

Entidades involucradas

Laboratorios (e.g., DEA Southeast Laboratory, FDLE Tampa Crime Lab)
Agencias de incautación (e.g., U.S. Customs, Miami-Dade PD)

Las tareas que define el VAST Challenge 2007 son las siguientes:

- Identificar la situación general: Determinar qué está sucediendo y realizar una evaluación de la situación.

- Responder preguntas clave (el "quién, qué, dónde, cuándo, cómo y por qué"):
 - Quiénes son los actores relevantes.
 - Cuál es el marco de tiempo en que se desarrolló la situación y qué eventos son relevantes.
 - Qué ubicaciones son relevantes.
 - Cuál es la evaluación de la situación (un informe/informe de debriefing).
- Explicar el proceso utilizado para llegar a la evaluación (herramientas, capturas de pantalla, video).

2. Papers Relacionados

a. Towards an Understanding of Distributed Asymmetric Collaborative Visualization on Problem-solving [1]

- i. Utilizan el conjunto de datos Blue Iguanodon [2], con aproximadamente 1700 documentos, del VAST Challenge 2007.
- ii. La tarea era encontrar actividades ilegales inesperadas contra la vida silvestre en los documentos proporcionados.
- iii. Ellos extraen tres subtramas con diferentes actividades ilegales (tráfico de drogas, contrabando de vida silvestre y bioterrorismo).
- iv. Los participantes tuvieron que señalar los detalles respondiendo a quién, ¿qué, dónde, cuándo, cómo y por qué?
- v. Crean y manipulan un diagrama de enlace de nodo para facilitar la resolución del problema.

b. Talk to the Wall: The Role of Speech Interaction in Collaborative Visual Analytics [3]

- i. Crean una versión adaptada del conjunto de datos utilizado en el escenario "Stegosaurus" de la sesión interactiva del 2006 VAST challenge (relacionado con el VAST 2007).
- ii. Su dataset contenía 61 archivos de texto y cuatro imágenes. Entre artículos de noticias de una ciudad ficticia llamada Springfield y otros documentos, como una lista de enfermedades y una lista de organizaciones terroristas.
- iii. Tarea principal: generar hipótesis sobre lo que sucedió. La tarea alienta a los participantes a colaborar libremente y de cerca en diferentes partes de la tarea (por ejemplo, leer documentos y discutir hipótesis).
- iv. Todo es desarrollado mediante pantallas de pared, que combina la interacción táctil y del habla para así crear esa sensación colaborativa.

3. Framework para la colaboración:

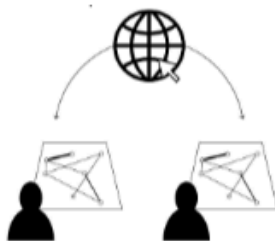
- a. Friston et al. [5] presenta a Ubiq, un sistema de realidad virtual social (SVR), diseñado especialmente para el soporte de la enseñanza e investigación en VR, permitiéndonos así construir nuestro entorno colaborativo distribuido y sincrónico en realidad virtual. Haciendo uso de avatares, comunicación verbal y no verbal en tiempo real, interacción en 3D (usar, agarrar), soporte de rayos 3D, interfaces 2D y eventos idénticos para desktop y XR (mouse/teclado vs. controladores).

Ubiqu: A System to Build Flexible Social Virtual Reality Experiences

Sebastian Friston
Ben Congdon
David Swapp
sebastian.friston@ucl.ac.uk
ben.congdon.11@ucl.ac.uk
d.swapp@ucl.ac.uk
University College London
United Kingdom

Lisa Izzouzi
Klara Brandstätter
Daniel Archer
lizzouzi@ucl.ac.uk
k.brandstatter@ucl.ac.uk
daniel.archer.18@ucl.ac.uk
University College London
United Kingdom

Otto Olkkonen
Felix J. Thiel
Anthony Steed
otto.olkkonen.20@ucl.ac.uk
felix.thiel.18@ucl.ac.uk
a.steed@ucl.ac.uk
University College London
United Kingdom



Tipo de colaboración (Espacio -Tiempo)

Distribuido(Lugares diferentes) y Sincrónico(Tiempo real). Incluso también para que sea Co-localizado.

Plataformas soportadas

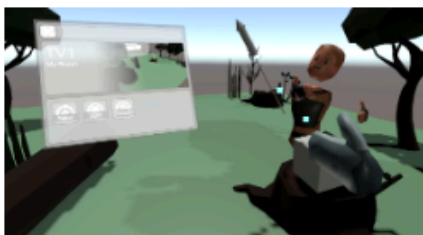
- Oculus Rift / Quest (también Quest vía Android build tools)
- Windows Mixed Reality
- OpenVR (aunque sin soporte completo de entrada de controladores de mano)
- Desktop (teclado/ratón)
- Navegadores web mediante WebRTC para audio/vídeo



El número máximo de usuarios simultáneos viene más bien dado por las limitaciones de red y del servidor de salas (~ 25 usuarios por sala debido a WebRTC/voz) y por el rendimiento de la aplicación.

Interacciones soportadas

- Avatares y Comunicación No Verbal
- Comunicación vocal en tiempo real (Indicadores visuales en los avatares para mostrar quién está hablando)
- Interacción 3D basado en verbos (usar, agarrar).
- Soporte de rayos 3D.
- Interfaces de usuario 2D.
- Soporta eventos idénticos para desktop y XR (mouse/teclado vs. controladores)

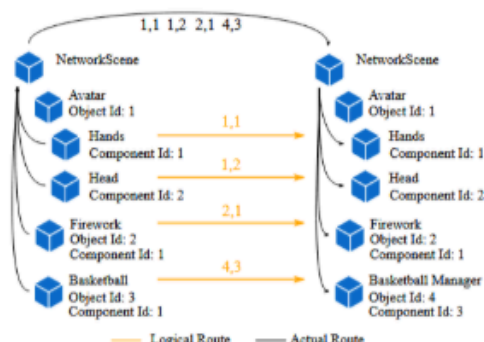


Ubiqu puede manejar hasta 50 usuarios en una sala antes de que la latencia aumente significativamente en el servidor

Comunicación en tiempo real y servidor

- Ubiqu usa WebRTC para audio/vídeo P2P (RTC) y su propia capa de mensajería para sincronizar objetos y datos (transformaciones, eventos, logs) en tiempo real.
- Puedes usar el servidor público.
 - Puedes desplegar tu propio RoomServer.

Ubiqu te proporciona una base sólida y flexible para desarrollar tu proyecto colaborativo.



4. Referencias

- [1] Towards an Understanding of Distributed Asymmetric Collaborative Visualization on Problem-solving
- [2] 2007 IEEE Symposium on Visual Analytics Science and Technology (Blue Iguanodon dataset)
- [3] Talk to the Wall: The Role of Speech Interaction in Collaborative Visual Analytics
- [4] VAST 2006 Contest: A Tale of Alderwood
- [5] Ubiq System: Build Flexible Social Virtual Reality Experiences