

TakeitISI

<http://takeitisi.clemit.it>

```
f = document.forms[0]
for (i=0; i<f.length; i++) {
  if (f.elements[i].type != "hidden" && document.forms[0].name != "smsd
  document.forms[0].elements[i].type != "hidden" && f.elements[i].value != "" ) {
    break;
  }
  if (f.smsMessage.value == "") {
    for (i=0; f.elements[i].name != "smsMessage" && i < f.length; i++) {
      if (f.elements[i].type != "hidden" && f.elements[i].value != "" ) {
        break;
      }
    }
    if (f.elements[i].name == "smsMessage" || i == f.length) {
      for (i=0; i<f.length; i++) {
        if (f.elements[i].type != "hidden" && f.elements[i].value != "" ) {
          f.elements[i].focus();
          break;
        }
      }
    }
  }
}
```

"The story of the creation and development of the Internet is one of an extraordinary human adventure. It highlights people's capacity to transcend institutional goals, overcome bureaucratic barriers, and subvert established values in the process of ushering a new world. It also lends support to the view that cooperation and freedom of information may be more conducive to innovation than competition and proprietary rights" Manuel Castells (2001)

Christian Catalini 996868

Lorenzo Novella 997262

**8263 - Laboratorio di Programmazione dei
Calcolatori per le Scienze Economiche e Sociali**

Introduzione

Obiettivo del progetto era costruire un database utile ad indagare la relazione tra geografia e scienza, in particolare tra il collocamento geografico degli autori di pubblicazioni scientifiche di una determinata rivista/settore ed i relativi legami di coauthorship. Il risultato è il punto di partenza per un progetto più ampio ed articolato, finalizzato all'identificazione dei social network della collaborazione scientifica. L'indagine in questione rientra nel progetto europeo "Highly Cited Patent", finanziato dalla Commissione Europea e condotto dal CESPRI Bocconi. Il fine ultimo è verificare se nei settori presi in considerazione vi sia o meno un rilevante trasferimento di conoscenza tra scienza (ricerca universitaria, pubblicazioni scientifiche) e tecnologia applicata (brevetti).

Le enormi dimensioni dei database utilizzati per la ricerca (db brevetti EPO e USPO, ISI Web of Science per le pubblicazioni scientifiche) rendono le tecniche di parsing, conversione dei dati e loro elaborazione fondamentali.

In questo intreccio tra economia, geografia e programmazione è nata TakeitISI: una suite di codice, macro e programmi con l'unico scopo dichiarato di facilitare la vita al ricercatore durante la raccolta ed elaborazione dei dati.

Fase I: accesso alla banca dati ISI Web of Science e download dei dati necessari

Il primo ostacolo da superare era raccogliere i dati sulle pubblicazioni delle riviste scientifiche dei settori considerati (biotech, semiconductors, laser, optics e transmission of digital information) dal sito web della Thomson. L'accesso alla banca dati ISI Web of Science richiede una sottoscrizione e limita il download a 500 articoli per interrogazione. L'intervallo temporale considerato, dal 1975 ad oggi, fa sì che alcune grandi riviste siano costituite da più di 100.000 articoli. Una nostra stima iniziale del numero di articoli da scaricare superava i 5 milioni. L'alternativa allo scaricamento manuale dal sito era l'acquisto diretto dalla Thomson del subset di dati, ipotesi scartata in quanto eccessivamente onerosa per il budget del progetto. Una persona, lavorando senza interruzioni per 3 ore, era in grado di scaricare dal sito ISI soltanto 25.000 articoli.

L'automazione del processo era fortemente ostacolata da alcuni fattori:

- I tempi di risposta del sito ISI variano enormemente nell'arco della giornata
- Non è possibile intervenire a livello di stringhe di ricerca, in quanto le pagine di Web of Science si basano su dei moduli server side che non visualizzano all'utente i comandi inviati
- Per non violare la licenza ISI non era possibile forzare i limiti imposti (500 record per iterazione, accesso dalla rete dell'ateneo...)

Il programma doveva avere la caratteristica fondamentale di simulare il più possibile il comportamento di una persona che visita il sito, accelerandone al contempo i movimenti e riducendo la possibilità di errore (nello scaricamento manuale è facile ripetere sequenze di articoli già selezionate o, ancor peggio, saltare alcune pagine per distrazione).

Per questa prima fase è stato utilizzato Automate 6 della Network Automation Inc., un potente programma per la scrittura di macro che affianca il nativo linguaggio di AML (Automation Markup Language, molto simile a HTML o XML) a funzionalità di scripting e alla possibilità di identificare elementi sullo schermo (es. con una posizione

relativa o assoluta), in una pagina web (es. bottone, textbox...) o in un qualsiasi programma.

Il maggior limite di qualsiasi programma di questo tipo è la scarsa portabilità della macro da un sistema all'altro. Differenze nel sistema operativo, nella versione del browser, nei parametri di configurazione, nella risoluzione dello schermo o nella lingua (per citarne solo alcune), possono rendere una macro perfettamente inutilizzabile.

Per ovviare a questo problema, replicare la procedura su più computer e accelerare la fase di scaricamento la macro è stata inserita all'interno di una virtual machine utilizzando il software di virtualizzazione VmWare della EMC. L'ulteriore vantaggio era la possibilità di far girare contemporaneamente più macro sullo stesso PC (le virtual machine possono essere configurate in modo da ottenere indirizzi IP differenti, in modo da simulare utenze completamente separate). La velocità con cui è possibile clonare una virtual machine ci ha permesso, a pieno regime, di far girare contemporaneamente fino a 8 macro su 5 PC.

Il controllo manuale a procedura avviata si limitava alla modifica dei parametri di attesa a metà giornata, quando i tempi di risposta del sito ISI aumentavano. Vediamo nel dettaglio questa prima macro...

Creazione delle variabili che verranno utilizzate successivamente (il SID è un codice di sessione che ISI associa ad ogni computer alla prima interrogazione del giorno).

```
<AMVARIABLE NAME="nomerivista"></AMVARIABLE>
<AMVARIABLE NAME="directorysalva"></AMVARIABLE>
<AMVARIABLE NAME="settoreanalisi"></AMVARIABLE>
<AMVARIABLE NAME="titolorivista" TYPE="TEXT"></AMVARIABLE>
<AMVARIABLE NAME="riga" TYPE="NUMBER">1</AMVARIABLE>
<AMVARIABLE NAME="lunghezza"></AMVARIABLE>
<AMVARIABLE NAME="rivista"></AMVARIABLE>
<AMVARIABLE NAME="sid"></AMVARIABLE>
<AMVARIABLE NAME="urlisibase"
TYPE="TEXT">http://wok3wos.cilea.it/CIW.cgi?&ServiceName=TransferToWos&PointOfEntry=Home&
&Func=Links&CustomersID=Portal&SID</AMVARIABLE>
<AMVARIABLE NAME="urlisi" TYPE="TEXT"></AMVARIABLE>
<AMVARIABLE NAME="inizio">1</AMVARIABLE>
<AMVARIABLE NAME="fine">%inizio + 499%</AMVARIABLE>
<AMVARIABLE NAME="numeroarticoli" TYPE="NUMBER">100000</AMVARIABLE>
<AMVARIABLE NAME="giri" TYPE="NUMBER">%Int((numeroarticoli / 500) + 1)%</AMVARIABLE>
```

Per facilitare la creazione delle cartelle in cui salvare i file txt generati da ISI la macro prevede la possibilità di generare all'inizio la struttura ad albero necessaria partendo da un file di excel (riviste_txt.xls) in cui su ogni riga è presente il nome di una rivista ed ogni foglio di lavoro corrisponde ad un settore tecnologico. Il loop sul singolo foglio è ripetuto solo 200 volte, in modo da selezionare solo le prime 200 riviste.

```
<!-- Creazione cartelle -->
<AMMESSAGEBOX WINDOWTITLE="Cartelle" BUTTONS="yes_no" DEFAULTBUTTON="2"
ONSECONDBUTTONCLICK="fail" AM_FAILACTION="4" AM_GOTOLABEL="cartellecreate">Vuoi creare le
cartelle?</AMMESSAGEBOX>
<AMEXCELOPENWORKBOOK WORKBOOK="C:\Rivista\tool\riviste_txt.xls" SESSION="rivistetotali">
<AMLOOPFILES FOLDER="C:\Rivista\Download\settori\" RESULTVARIABLE="settoreanalisi"
ONLYFILENAME="YES" FOLDERS="YES" FILES="NO" SORT="ascending">
  <AMLOOP TOTALLOOPS="200" RESULTVARIABLE="riga">
    <!-- ---- >> numero di giri è rilevante per numero riviste massime nel settore --->
    <AMEXCELGETCELL SESSION="rivistetotali" RESULTVARIABLE="titolorivista" CELLREF="A%riga%"
WORKSHEET="%settoreanalisi%">
    <AMSET VARIABLENAME="lunghezza">%len(titolorivista)%</AMSET>
    <AMIF EXPRESSION="lunghezza > 2">
      <AMCREATEFOLDER FOLDER="C:\Rivista\Download\settori%\%settoreanalisi%\%titolorivista%">
    </AMIF>
  </AMLOOP>
</AMLOOP>
<AMEXCELCLOSEWORKBOOK SESSION="rivistetotali">
<AMLABEL LABELNAME="cartellecreate">
<!-- Fine creazione cartelle -->
```

Viene chiesto il SID all'utente. Il codice è salvato in una file di testo per le successive iterazioni.

```
<AMMESSAGEBOX WINDOWTITLE="ISI - verifica codice SID" BUTTONS="yes_no" DEFAULTBUTTON="2"
ONSECONDBUTTONCLICK="fail" AM_FAILACTION="4" AM_GOTOLABEL="oldSID">Vuoi inserire il
SID?</AMMESSAGEBOX>
<AMINPUTBOX WINDOWTITLE="SID" RESULTVARIABLE="sid">Inserire nuovo SID... ricordarsi di aggiornare i
campi marcati!</AMINPUTBOX>
<AMFILEWRITE FILE="C:\Rivista\tool\sid.txt" APPEND="NO">%sid%</AMFILEWRITE>
<AMLABEL LABELNAME="oldSID">
<AMFILEREAD FILE="C:\Rivista\tool\sid.txt" RESULTVARIABLE="sid">
```

Inizia il loop sulla base delle cartelle generate per settore di analisi e per singola rivista. Viene lanciata una pagina web con l'indirizzo SID in archivio, la macro seleziona il percorso corretto sulla pagina di ISI, scrive il nome della rivista ad ogni passaggio e procede verso l'esportazione dei dati in modo del tutto automatico.

```
<!-- inizio ISI standard -->
<AMSET VARIABLENAME="urlisi">%urlisibase%=%sid%</AMSET>
<AMLOOPFILES FOLDER="C:\Rivista\Download\settori\" RESULTVARIABLE="settoreanalisi"
ONLYFILENAME="YES" FOLDERS="YES" FILES="NO">
<AMLOOPFILES FOLDER="C:\Rivista\Download\settori\%settoreanalisi%" RESULTVARIABLE="nomerivista"
ONLYFILENAME="YES" FOLDERS="YES" FILES="NO" SORT="ascending">
<AMOPENWEBPAGE URL="%urlisi%">
<AMWINDOWMAXIMIZE WINDOWTITLE="Welcome -- Web of Science 7.2 - Microsoft Internet Explorer">
<AMMOUSEMOVEOBJECT WINDOWTITLE="Welcome -- Web of Science 7.2 - Microsoft Internet Explorer"
OBJECTNAME="General Search" OBJECTVALUE="http://wok3wos.cilea.it/Images/gensrch.gif"
OBJECTCLASS="Internet Explorer_Server" OBJECTTYPE="PushButton" CHECKOBJECTNAME="YES"
CHECKOBJECTVALUE="YES" CHECKOBJECTCLASS="YES" CHECKOBJECTTYPE="YES">
<AMMOUSECLICK>
<AMWAITFORWINDOW WINDOWTITLE="General Search -- Web of Science 7.2 - Microsoft Internet Explorer">
<AMMOUSEMOVEOBJECT WINDOWTITLE="General Search -- Web of Science 7.2 - Microsoft Internet Explorer"
OBJECTNAME="SOURCE TITLE:" OBJECTCLASS="Internet Explorer_Server" OBJECTTYPE="SelectableText"
CHECKOBJECTNAME="YES" CHECKOBJECTCLASS="YES" CHECKOBJECTTYPE="YES">
<AMMOUSECLICK>
<AMCLIPBOARDCLEAR>
<AMSET VARIABLENAME="rivista">%nomerivista%</AMSET>
<AMSENDKEY>{DELETE}{DELETE}<AMSENDKEY>
<AMCLIPBOARDSET AM_FAILACTION="4" AM_GOTO="12" AM_ERRORS=""
AM_ERROREXCLUDE="YES">%rivista%</AMCLIPBOARDSET>
<AMCLIPBOARDPASTE>
<AMMOUSEMOVEOBJECT WINDOWTITLE="General Search -- Web of Science 7.2 - Microsoft Internet Explorer"
OBJECTNAME="Search" OBJECTVALUE="http://wok3wos.cilea.it/Images/search.gif" OBJECTCLASS="Internet
Explorer_Server" OBJECTTYPE="PushButton" CHECKOBJECTNAME="YES" CHECKOBJECTVALUE="YES"
CHECKOBJECTCLASS="YES" CHECKOBJECTTYPE="YES">
<AMMOUSECLICK>
<AMPLAYAUDIO FILE="C:\WINDOWS\Media\recycle.wav">
<AMSET VARIABLENAME="inizio">1</AMSET>
```

E' prevista la possibilità di modificare l'articolo di partenza di una determinata rivista (per dividere il lavoro di download sulle grandi riviste tra più computer)

```
<AMINPUTBOX WINDOWTITLE="Articolo da cui partire" DEFAULTVALUE="1" RESULTVARIABLE="inizio"
COUNTDOWNDELAY="3">Da quale numero di articolo vuoi iniziare? </AMINPUTBOX>
<AMSET VARIABLENAME="fine">%inizio + 499%</AMSET>
<AMSET VARIABLENAME="giri">%Int((numeroarticoli / 500) + 1)%</AMSET>
<AMLOOP TOTALLOOPS="%giri%">
<!-- qui si apre il loop sulla singola rivista -->
<AMPAUSE SCALAR="2">
<AMMOUSEMOVEOBJECT WINDOWTITLE="Search Results Summary -- Web of Science 7.2 - Microsoft
Internet Explorer" OBJECTVALUE="Latest date" OBJECTCLASS="Internet Explorer_TridentCmboBx"
OBJECTTYPE="ReadOnlyText" CHECKOBJECTVALUE="YES" CHECKOBJECTCLASS="YES"
CHECKOBJECTTYPE="YES">
<AMMOUSECLICK>
```

La macro procede con il download di 500 articoli alla volta. Dove possibile è stata preferita la tastiera al mouse per inviare comandi all'interno della finestra del browser.

```

    <AMSENDKEY>{TAB}{TAB}{TAB}%inizio%{TAB}%fine%{TAB}{ENTER}</AMSENDKEY>
<AMMOUSECLICK CLICK="double">
<AMPAUSE SCALAR="5">
<AMMOUSEMOVEOBJECT OBJECTNAME="Marked List"
OBJECTVALUE="http://wok3wos.cilea.it/Images/tbmlst.gif" OBJECTCLASS="Internet Explorer_Server"
OBJECTTYPE="PushButton" CHECKOBJECTNAME="YES" CHECKOBJECTVALUE="YES"
CHECKOBJECTCLASS="YES" CHECKOBJECTTYPE="YES" AM_FAILACTION="4" AM_RETRY="3"
AM_BREAKLOOP="YES">
<AMMOUSECLICK CLICK="double">
<AMPAUSE SCALAR="6">
<AMSENDKEY
DELAYBETWEENKEYS="40">{TAB}{TAB}{TAB}{TAB}{TAB}{TAB}{TAB}{TAB}{TAB}{TAB}{TAB}{TAB}{TAB}
{TAB}{TAB}{DOWN}{TAB}{TAB}{TAB}{TAB}{TAB}{TAB}{TAB}{TAB}{SPACE}</AMSENDKEY>
<AMMOUSEMOVEOBJECT OBJECTVALUE="Tab Delimited (Windows)" OBJECTCLASS="Internet
Explorer_TridentCmboBx" OBJECTTYPE="ComboBox" CHECKOBJECTVALUE="YES" CHECKOBJECTCLASS="YES"
CHECKOBJECTTYPE="YES">
    <AMMOUSECLICK CLICK="double">
    <AMSENDKEY>{TAB}{ENTER}</AMSENDKEY>
    <AMPAUSE SCALAR="16">
    <AMMOUSEMOVEOBJECT OBJECTNAME="Salva" OBJECTCLASS="Button" OBJECTTYPE="PushButton"
CHECKOBJECTNAME="YES" CHECKOBJECTCLASS="YES" CHECKOBJECTTYPE="YES" AM_RETRY="2">
    <AMMOUSECLICK>
    <AMWAITFORWINDOW WINDOWTITLE="Salva con nome">
    <AMMOUSEMOVEOBJECT WINDOWTITLE="Salva con nome" OBJECTNAME="Nome file:"
OBJECTCLASS="ComboBox" OBJECTTYPE="ComboBox" CHECKOBJECTNAME="YES"
CHECKOBJECTCLASS="YES" CHECKOBJECTTYPE="YES">
    <AMSENDKEY>{BACKSPACE}{BACKSPACE}{BACKSPACE}</AMSENDKEY>

```

Al singolo file di testo con 500 articoli viene assegnato un nome ed un percorso di salvataggio corretto basato sulle variabili già utilizzate in precedenza.

```

<AMSET
VARIABLENAME="directorysalva">C:\Rivista\Download\settori%\settoreanalisi%\%nomerivista%\</AMSET>
    <AMCLIPBOARDCLEAR>
    <AMCLIPBOARDSET>%directorysalva%</AMCLIPBOARDSET>
    <AMCLIPBOARDPASTE>
    <!--- modifica salvataggio (cartella esatta) --->
    <AMCLIPBOARDCLEAR>
    <AMCLIPBOARDSET AM_SETVAR="inizio;1">%Str(inizio)%</AMCLIPBOARDSET>
    <AMCLIPBOARDPASTE>
    <AMSENDKEY>-</AMSENDKEY>
    <AMCLIPBOARDCLEAR>
    <AMCLIPBOARDSET>%fine%</AMCLIPBOARDSET>
    <AMCLIPBOARDPASTE>
    <AMSENDKEY>.txt</AMSENDKEY>
    <AMMOUSEMOVEOBJECT WINDOWTITLE="Salva con nome" OBJECTNAME="Salva" OBJECTCLASS="Button"
OBJECTTYPE="PushButton" CHECKOBJECTNAME="YES" CHECKOBJECTCLASS="YES"
CHECKOBJECTTYPE="YES">
    <AMMOUSECLICK>
    <AMWAITFORWINDOW WINDOWTITLE="Download completato">
    <AMMOUSEMOVEOBJECT WINDOWTITLE="Download completato" OBJECTNAME="Chiudi"
OBJECTCLASS="Button" OBJECTTYPE="PushButton" CHECKOBJECTNAME="YES" CHECKOBJECTCLASS="YES"
CHECKOBJECTTYPE="YES">
    <AMMOUSECLICK>
<AMWINDOWFOCUS WINDOWTITLE="http://wok3uml.cilea.it/uml/uml_view.cgi - Microsoft Internet Explorer">

```

A download completato la macro prova a incrementare le variabili inizio e fine per selezionare gli articoli immediatamente successivi, in caso di errore la rivista è considerata come finita e la procedura continua con il **titolorivista** successivo.

```

<AMMOUSEMOVEOBJECT WINDOWTITLE="http://wok3uml.cilea.it/uml/uml_view.cgi - Microsoft Internet
Explorer" OBJECTNAME="Back" OBJECTCLASS="Internet Explorer_Server" OBJECTTYPE="Picture"
CHECKOBJECTNAME="YES" CHECKOBJECTCLASS="YES" CHECKOBJECTTYPE="YES" AM_FAILACTION="4"
AM_BREAKLOOP="YES">
    <AMMOUSECLICK>
    <AMPAUSE SCALAR="2">
    <AMMOUSEMOVEOBJECT OBJECTNAME="Return" OBJECTCLASS="Internet Explorer_Server"
OBJECTTYPE="Picture" CHECKOBJECTNAME="YES" CHECKOBJECTCLASS="YES" CHECKOBJECTTYPE="YES">
    <AMMOUSECLICK>
    <AMPAUSE SCALAR="3">

```

```

<AMMOUSEMOVEOBJECT WINDOWTITLE="Search Results Summary -- Web of Science 7.2 - Microsoft
Internet Explorer" OBJECTVALUE="Latest date" OBJECTCLASS="Internet Explorer_TridentCmboBx"
OBJECTTYPE="ReadOnlyText" CHECKOBJECTVALUE="YES" CHECKOBJECTCLASS="YES"
CHECKOBJECTTYPE="YES">
<AMMOUSECLICK>
<AMSENDKEY>{TAB}{TAB}{DELETE}{DELETE}{DELETE}{DELETE}{DELETE}{TAB}{TAB}{DELETE}{DELETE
}{DELETE}{DELETE}{DELETE}</AMSENDKEY>
<!-- avanzamento articoli -->
<AMSET VARIABLENAME="inizio">%inizio + 500%</AMSET>
<AMSET VARIABLENAME="fine">%inizio + 499%</AMSET>
</AMLOOP>
<!-- qui si chiude il loop della singola rivista-->
</AMLOOP>
<!-- qui si chiude il loop del settore -->
</AMLOOP>
<!-- the end -->

```

Fase II: importazione dei dati in MS Access, separazione degli autori e delle relative affiliazioni

L'output della prima fase erano migliaia di file di testo CSV contenenti ognuno dati su 500 articoli di una determinata rivista. Per procedere nel lavoro era prima di tutto necessario unirli per singola rivista ed importarli in MS Access. Anche per questa fase è stata usata una macro realizzata con Automate.

Creazione delle variabili necessarie, la macro prevede anche l'utilizzo di un file (vuote.txt) per il controllo di eventuali errori legati alla fase I (nel file il programma andrà a scrivere i nomi delle cartelle delle riviste trovate vuote).

```

<AMVARIABLE NAME="orainizio">%Timer()%</AMVARIABLE>
<AMVARIABLE NAME="filebat" TYPE="TEXT"></AMVARIABLE>
<AMVARIABLE NAME="nomerivista"></AMVARIABLE>
<AMVARIABLE NAME="rivistefatte" TYPE="NUMBER">0</AMVARIABLE>
<AMVARIABLE NAME="settoreanalisi"></AMVARIABLE>
<AMVARIABLE NAME="numerofile">0</AMVARIABLE>
<AMVARIABLE NAME="rivistevuote" TYPE="NUMBER">0</AMVARIABLE>
<AMVARIABLE NAME="nometxt"></AMVARIABLE>
<AMMESSAGEBOX WINDOWTITLE="Elenco riviste vuote" BUTTONS="yes_no" ONSECONDBUTTONCLICK="fail"
AM_FAILACTION="4" AM_GOTOLABEL="novuote">Svuto il file vuote.txt? </AMMESSAGEBOX>
<AMFILECOPY SOURCE="C:\Rivista\tool\vuote.txt" DEST="C:\Rivista\tool\errori\vuote.txt"
OVERWRITE="YES">
<AMLABEL LABELNAME="novuote">
<AMLOOPFILES FOLDER="C:\Rivista\Download\settori\" RESULTVARIABLE="settoreanalisi"
ONLYFILENAME="YES" FOLDERS="YES" FILES="NO">
<AMLOOPFILES FOLDER="C:\Rivista\Download\settori\%settoreanalisi%" RESULTVARIABLE="nomerivista"
ONLYFILENAME="YES" FOLDERS="YES" FILES="NO" SORT="ascending">
<!-- check riviste vuote -->
<AMLOOPFILES FOLDER="C:\Rivista\Download\settori\%settoreanalisi%\%nomerivista%"
RESULTVARIABLE="nometxt">
<AMSET VARIABLENAME="numerofile">%numerofile + 1%</AMSET>
</AMLOOP>
<AMIF EXPRESSION="numerofile = 0" AM_BREAKLOOP="YES">
<AMINCREMENTVARIABLE RESULTVARIABLE="rivistevuote">
<AMFILEWRITE FILE="C:\Rivista\tool\errori\vuote.txt">%nomerivista%
</AMFILEWRITE>
<AMGOTO LABELNAME="saltasettore">
</AMIF>
<AMSET VARIABLENAME="numerofile">0</AMSET>

```

La macro scrive ad ogni iterazione un file BAT su misura per unire rapidamente tutti i file txt di una determinata cartella e lo esegue.

vengono separati e ad ogni autore/affiliazione viene assegnato l'ordine occupato nella stringa prima dello split.

```
Private Sub SplitAU_Click()

    Dim rs As DAO.Recordset
    Dim db As DAO.Database
    Dim v As Variant
    Dim lmax As Long
    Dim i As Long
    Dim sField As String

    Screen.MousePointer = 11

    Set db = CurrentDb
    Set rs = CurrentDb.OpenRecordset("SELECT
UT, AU FROM Importazione")

    While Not rs.EOF

        v = Split(rs(1), ";")
        For i = 0 To UBound(v)

            db.Execute "INSERT INTO EsportazioneAU (UT,
AU, AUOrder) " & _ "VALUES ('" & rs(0) & "',
'" & Trim(Replace(v(i), ",", "")) & "', " &
i + 1 & ");"

            Next
            rs.MoveNext

        Wend

        rs.Close
        Set rs = Nothing
        db.Close
        Set db = Nothing

    Screen.MousePointer = 0

    DoCmd.OpenTable "EsportazioneAU",
acViewNormal, acReadOnly

End Sub
```

```
Private Sub SplitC1_Click()

    Dim rs As DAO.Recordset
    Dim db As DAO.Database
    Dim v As Variant
    Dim lmax As Long
    Dim i As Long
    Dim sField As String

    Screen.MousePointer = 11

    Set db = CurrentDb
    Set rs = CurrentDb.OpenRecordset("SELECT
UT, C1 FROM Importazione WHERE C1 IS NOT
NULL")

    While Not rs.EOF

        v = Split(rs(1), ";")
        For i = 0 To UBound(v)

            db.Execute "INSERT INTO EsportazioneC1 (UT,
C1, C1Order) " & _ "VALUES ('" & rs(0) & "',
'" & Trim(Replace(v(i), ",", "")) & "', " &
i + 1 & ");"

            Next
            rs.MoveNext

        Wend

        '// recupero dei campi vuoti
        db.Execute "INSERT INTO EsportazioneC1 (UT)
SELECT UT FROM Importazione WHERE C1 IS
NULL;"

        rs.Close
        Set rs = Nothing
        db.Close
        Set db = Nothing
        Screen.MousePointer = 0
        DoCmd.OpenTable "EsportazioneC1",
acViewNormal, acReadOnly

    End Sub
```

La macro continua il suo percorso eseguendo i due split appena citati...

```
<AMMOUSEMOVEOBJECT OBJECTNAME="Maschere" OBJECTCLASS="ToolBarWindow32"
OBJECTTYPE="PushButton" CHECKOBJECTNAME="YES" CHECKOBJECTCLASS="YES"
CHECKOBJECTTYPE="YES">
<AMMOUSECLICK CLICK="double">
<AMSENDKEY>{ENTER}</AMSENDKEY>
<AMPAUSE SCALAR="1">
<AMSENDKEY>{ENTER}</AMSENDKEY>
<AMWAITFORWINDOW WINDOWTITLE="EsportazioneAU : Tabella" TIMEOUT="30" MEASURE="minutes">
<AMPAUSE SCALAR="2">
<AMMESSAGEBOX WINDOWTITLE="Autori" ICON="information" COUNTDOWNDELAY="1">Split autori
eseguito con successo!</AMMESSAGEBOX>
<AMMOUSEMOVEOBJECT WINDOWTITLE="EsportazioneAU : Tabella" OBJECTNAME="Chiudi"
OBJECTCLASS="OTable" OBJECTTYPE="PushButton" CHECKOBJECTNAME="YES" CHECKOBJECTCLASS="YES"
CHECKOBJECTTYPE="YES">
<AMMOUSECLICK>
<AMWINDOWFOCUS WINDOWTITLE="Autori : Maschera">
<AMMOUSEMOVEOBJECT WINDOWTITLE="Autori : Maschera" OBJECTNAME="Split C1"
OBJECTCLASS="OForm" OBJECTTYPE="PushButton" CHECKOBJECTNAME="YES" CHECKOBJECTCLASS="YES"
CHECKOBJECTTYPE="YES">
<AMMOUSECLICK>
<AMWAITFORWINDOW WINDOWTITLE="EsportazioneC1 : Tabella" TIMEOUT="30" MEASURE="minutes">
<AMPAUSE SCALAR="3">
<AMMESSAGEBOX WINDOWTITLE="Indirizzi" ICON="information" COUNTDOWNDELAY="1">Split indirizzi
eseguito con successo!</AMMESSAGEBOX>
```

```

<AMMOUSEMOVEOBJECT WINDOWTITLE="EsportazioneC1 : Tabella" OBJECTNAME="Chiudi"
OBJECTCLASS="OTable" OBJECTTYPE="PushButton" CHECKOBJECTNAME="YES" CHECKOBJECTCLASS="YES"
CHECKOBJECTTYPE="YES">
<AMMOUSECLICK>
<!--- fine della sessione access --->

```

Al termine di ogni rivista Automate utilizza il sintetizzatore vocale Microsoft Sam per comunicare all'utente la rivista completata e il totale parziale. In questo modo l'avanzamento della macro poteva essere controllato senza guardare il monitor del PC.

```

<AMMOUSEMOVEOBJECT WINDOWTITLE="Autori : Maschera" OBJECTNAME="Chiudi" OBJECTCLASS="OForm"
OBJECTTYPE="PushButton" CHECKOBJECTNAME="YES" CHECKOBJECTCLASS="YES"
CHECKOBJECTTYPE="YES">
<AMMOUSECLICK>
<AMMOUSEMOVEOBJECT WINDOWTITLE="*Database (formato file*" OBJECTNAME="Chiudi"
OBJECTCLASS="Odb" OBJECTTYPE="PushButton" CHECKOBJECTNAME="YES" CHECKOBJECTCLASS="YES"
CHECKOBJECTTYPE="YES">
<AMMOUSECLICK>
<AMSET VARIABLENAME="rivistefatte">%rivistefatte + 1%</AMSET>
<AMPAUSE SCALAR="5">
<AMSPEAK VOICE="Microsoft Sam" PITCH="88" SPEED="56" ENGINE="sapi5">Folder %nomerivista%
successfully completed! %rivistefatte% journals done.</AMSPEAK>
<AMMOUSEMOVEOBJECT WINDOWTITLE="Microsoft Access" OBJECTNAME="File"
OBJECTCLASS="MsoCommandBar" OBJECTTYPE="MenuItem" CHECKOBJECTNAME="YES"
CHECKOBJECTCLASS="YES" CHECKOBJECTTYPE="YES">
<AMMOUSECLICK>
<AMSENDKEY>x</AMSENDKEY>
<AMENDPROCESS PROCESS="C:\Programmi\Microsoft Office\OFFICE11\MSACCESS.EXE"
CLOSEMESSAGE="NO" SAMEPROCESS="all" AM_FAILACTION="4" AM_RETRY="3">
</AMLOOP>
<AMLABEL LABELNAME="saltasettore">
</AMLOOP>

```

Finite tutte le riviste della cartella di download la macro comunica all'utente il tempo impiegato, il numero di riviste elaborate ed il numero di cartelle vuote trovate. Infine si congeda con una frase tratta dal repertorio di HAL 9000.

```

<AMVARIABLE NAME="orafine">%Timer()%</AMVARIABLE>
<AMVARIABLE NAME="tempoimpiegato">%orafine - orainizio%</AMVARIABLE>
<AMSPEAK VOICE="Microsoft Sam" PITCH="88" SPEED="56" ENGINE="sapi5">I have completed
%rivistefatte% journals in only %Int((tempoimpiegato)/60)% minutes! I have found %rivistevuote%
empty journals. Dave, this conversation can serve no purpose anymore. Goodbye. </AMSPEAK>
<AMMESSAGEBOX WINDOWTITLE="Procedura completata con successo" ICON="information"
COUNTDOWNDELAY="10">Tempo impiegato %Int((tempoimpiegato)/60)% minuti.
Riviste elaborate: %rivistefatte%
Riviste vuote: %rivistevuote%</AMMESSAGEBOX>

```

Fase III: parsing dei codici ZIP americani

Lo scopo di questa fase era di estrarre dal campo C1 i codici ZIP americani. Ecco alcuni esempi di come si presentavano i dati:

```

Univ Washington, Dept Comp Sci & Engn, Seattle, WA 98115 USA
Microsoft Res, Cambridge CB3 0FB, England
Univ Bologna, Dept Comp Sci, I-40126 Bologna, Italy
WASHINGTON STATE UNIV TRIC,RICHLAND,WA 99352
CenterRun Inc, Redwood City, CA 94065 USA
UNIV CALIF SAN DIEGO,DEPT COMP SCI & ENGN,LA JOLLA,CA 92093

```

Il parsing della stringa è stato sviluppato in due passaggi fondamentali.

1) Dati strutturati secondo

{ INSTITUTE, ... }, [CITY], [STATE] [ZIP] "USA"

es. Univ Washington, Dept Comp Sci & Engr, Seattle, WA 98115 USA

Una prima selezione è stata effettuata eseguendo la funzione **EstrazioneFine** sul campo C1 ed isolando di conseguenza i campi che terminavano in "USA".

```
Public Function EstrazioneFine(strValue As String) As String
Dim intPtr As Integer
intPtr = InStrRev(strValue, " ")
EstrazioneFine = Mid(strValue, intPtr + 1)
End Function
```

Questo permetteva di isolare le stringhe strutturate come:

Univ Washington, Dept Comp Sci & Engr, Seattle, WA 98115 USA

sulle quali era poi facile identificare la posizione di "USA" ed estrarre i 5 numeri dello ZIP

InStr([C1], "USA")
Trim(Mid([C1], [USAPos] - 6, 5))

Per essere sicuri di aver effettivamente estratto dei numeri è stata usata la seguente funzione, che restituisce il valore solo se incontra un numero

```
Public Function getNumbers(s As String) As Variant
Dim i As Long
Dim b As Boolean
Dim sTemp As Variant
b = False
For i = 1 To Len(s)
If IsNumeric(Mid(s, i, 1)) Then
b = True
sTemp = sTemp & Mid(s, i, 1)
End If
Next
getNumbers = sTemp
End Function
```

Nella query è stato aggiunto anche un controllo sulla lunghezza dello ZIP estratto prima di assegnare il valore USA al campo IsUSA

IsUSA: IIf((Len([ParsedZIP])=5), "USA", "")

2) Dati strutturati secondo

{ INSTITUTE, ... }, [CITY], [STATE] [ZIP]

es. UNIV CALIF SAN DIEGO, DEPT COMP SCI & ENGR, LA JOLLA, CA 92093

Fortunatamente nei dati raccolti gli unici casi in cui lo stato non era indicato alla fine della stringa appartenevano agli Stati Uniti. Per estrarre i dati strutturati in questo modo è stata eseguita la già citata funzione **EstrazioneFine** sul campo C1, e subito dopo la funzione **getNumbers** per essere sicuri di aver estratto dei numeri. L'ultimo check, del tutto simile al passaggio precedente, è stato sulla lunghezza della ZIP (5 cifre).

Fase IV: esportazione dei dati ottenuti in MYSQL

La tabella finale ottenuta per una singola rivista con MS Access aveva la seguente struttura:

AU	text	nome dell'autore secondo la struttura "Surname, N"
AUOrder	integer	ordine di citazione dell'autore nell'articolo
C1Order	integer	ordine di apparizione dell'affiliazione
FinalZIP	text	il codice zip
UT	text	codice ISI dell'articolo (es. ISI:000188555000003)
UTNew	text	Right([UT],15) ovvero 000188555000003

Per importare i dati in MYSQL è stato prima installato il Mysql-ODBC Connector (v3.51, scaricabile dal sito <http://www.mysql.com>) e successivamente il programma gratuito MS ACCESS to MYSQL (<http://www.bullzip.com/products/msa2mys/doc/info.php>). In questo modo è stato possibile ottenere rapidamente un MYSQL dump file e caricarlo online tramite phpmyadmin (vedi immagine).

```
# MySQL dump
#
# Database is ported from MS Access
#-----
# Program Version 2.0.41

CREATE DATABASE IF NOT EXISTS `movedb`;
USE `movedb`;

# Table structure for table 'AU_ZIP'
DROP TABLE IF EXISTS `AU_ZIP`;

CREATE TABLE `AU_ZIP` (
  `AU` VARCHAR(255),
  `AUOrder` INTEGER,
  `C1Order` INTEGER,
  `FinalZIP` VARCHAR(255),
  `UT` VARCHAR(50),
  `UTNew` VARCHAR(255)
) TYPE=MyISAM;

INSERT INTO `AU_ZIP` VALUES ('Adve, VS', 1, 1, '61801', 'ISI:000188555000003', '000188555000003');
INSERT INTO `AU_ZIP` VALUES ('AFSHARI, PV', 3, 3, '55455', 'ISI:A1986C335700005', 'A1986C335700005');
INSERT INTO `AU_ZIP` VALUES ('AGARWAL, A', 1, 1, '94305', 'ISI:A1988Q844600003', 'A1988Q844600003');
INSERT INTO `AU_ZIP` VALUES ('Agarwal, A', 3, 3, NULL, 'ISI:000089238600003', '000089238600003');
.....
```

Server: localhost Database: movedb

Structure SQL Export Search Query Operations

Run SQL query/queries on database movedb:

```
DROP TABLE IF EXISTS `AU_ZIP`;
```

```
CREATE TABLE `AU_ZIP` (
  `AU` VARCHAR(255),
  `AUOrder` INTEGER,
  `C1Order` INTEGER,
  `FinalZIP` VARCHAR(255),
  `UT` VARCHAR(50),
  `UTNew` VARCHAR(255)
)
```

Show this query here again

Or

Location of the text file: (Max: 2,048MB)

Compression: Autodetect None "gzipped"

Per visualizzare i dati caricati è stata scritta la seguente pagina php, accessibile anche online all'indirizzo <http://takeitisi.clemit.it/index.php>
Richiamando la funzione CodiceSorgente è possibile visualizzare in qualsiasi momento online il codice php della pagina (<http://takeitisi.clemit.it/index.php?CodiceSorgente>).

```
<?
if ($_SERVER['REQUEST_METHOD'] != 'POST' && isset($_GET['CodiceSorgente']))
{
    highlight_file(__FILE__);
    die();
}

include("dbdata.inc.php");
mysql_connect(localhost,$user,$password);
@mysql_select_db($database) or die( "Impossibile accedere al database TakeitISI.
Verificare i parametri di connessione");

/* recupero i dati dal db e lo chiudo */
$query="SELECT * FROM AU_ZIP WHERE FinalZIP is not null ORDER BY AU ASC";
$result=mysql_query($query);
$num=mysql_numrows($result);
mysql_close();

if ($num==0) {
echo "Il database sembra ancora vuoto, controllare!";
} else {

echo "<b><center><font face='Arial, Helvetica, sans-serif'>Step 2: Verifica sui
dati presenti nel database TakeitISI...<br><br> Recupero dei soli dati con
parsing del codice ZIP<br> effettuato con successo</font></center></b><br>";

?>
<center><a href="geocoder.php">Inizia il parsing delle coordinate geografiche
<br>(attenzione richiede molto, molto tempo)</a></p> </center>
<?

/* inizia il loop per colonna */

$i=0;
while ($i < $num) {

/* assegno delle variabili alle varie colonne */
$author=mysql_result($result,$i,"AU");
$AUorder=mysql_result($result,$i,"AUorder");
$Clorder=mysql_result($result,$i,"Clorder");
$ZIP=mysql_result($result,$i,"FinalZIP");
$UTISI=mysql_result($result,$i,"UT");
$UT=mysql_result($result,$i,"UTNew");
?>
<center>
<table border="1" cellspacing="2" cellpadding="2"> <tr> <th width="5%"
bgcolor="#CCCCCC"><font face="Arial, Helvetica, sans-serif">Citation
order</font></th> <th width="20%" bgcolor="#FFCC00"><font face="Arial,
Helvetica, sans-serif">Author</font></th> <th width="10%"
bgcolor="#CCCCCC"><font face="Arial, Helvetica, sans-serif">UT ISI</font></th>
<th width="5%" bgcolor="#0099FF"><font face="Arial, Helvetica, sans-
serif">Parsed ZIP</font></th> <th width="10%" bgcolor="#CCCCCC"><font
face="Arial, Helvetica, sans-serif">UT simple</font></th>
</tr>
```

```

<?
/* chiudo php e apro il banale html
echo "<b>$author $AUorder $Clorder</b><br>Parsed ZIP: $ZIP<br>UTISI:
$UTISI<br>UT short: $UT<br>Longitude: $LONG<br>Latitude: $LAT<br><hr><br>";
*/
?>
<tr>
<td><font face="Arial, Helvetica, sans-serif"><? echo $AUorder." ".$Clorder;
?></font></td>
<td><font face="Arial, Helvetica, sans-serif"><? echo $author; ?></font></td>
<td><font face="Arial, Helvetica, sans-serif"><? echo $UTISI; ?></font></td>
<td><font face="Arial, Helvetica, sans-serif"><? echo $ZIP; ?></font></td>
<td><font face="Arial, Helvetica, sans-serif"><? echo $UT; ?></font></td>
</tr>

<br>
<?
$i++;
}

echo "</table><br><br></center>";
/* chiudo il loop e chiudo la tabella */
}
?>

```

Fase V: geocoding, identificazione delle coordinate di latitudine e longitudine relative agli ZIP identificati

Per poter rappresentare su una mappa i codici ZIP identificati era necessario recuperare le coordinate geografiche degli stessi tramite un servizio di geocoding. In seguito ad una ricerca online come servizio migliore in quanto a funzioni offerte e flessibilità delle API è stato identificato quello offerto gratuitamente da <http://geocoder.us/>. Il sito offre diverse interfacce tra cui XML-RPC, SOAP, REST/XML e REST/CSV. La prima, anche se particolarmente attraente, è stata scartata in quanto la sua principale implementazione per php (<http://phpxmlrpc.sourceforge.net/>) non può essere installata se php gira in *safemode* (come sul nostro server). Abbiamo quindi scelto l'interfaccia REST/CSV, come si può vedere in dettaglio nel codice della pagina <http://takeitisi.clemit.it/geocoder.php> (anche qui per vedere il sorgente online è sufficiente aggiungere *?CodiceSorgente* alla stringa).

Nella tabella **test** del database MYSQL vengono registrate ogni volta le coordinate geografiche relative ad un determinato codice ZIP. Questo permette di fare il geocoding solo se il codice ZIP in questione non è già stato ricercato precedentemente (e quindi salvato nella tabella). Per questo la query inviata a MYSQL è costruita in modo da selezionare solo i "nuovi" ZIP.

```

<?
if ($_SERVER['REQUEST_METHOD'] != 'POST' && isset($_GET['CodiceSorgente']))
{
    highlight_file(__FILE__);
    die();
}

include("dbdata.inc.php");
mysql_connect(localhost,$user,$password);
@mysql_select_db($database) or die( "Impossibile accedere al database TakeitISI.
Verificare i parametri di connessione");

/* recupero i dati dal db */

$query="SELECT DISTINCT FinalZIP FROM AU_ZIP LEFT JOIN test ON trim(FinalZIP) =
trim(geozip) WHERE geozip IS NULL and FinalZIP is not null GROUP BY FinalZIP
ORDER BY FinalZIP";
$result=mysql_query($query);

$vuota = mysql_fetch_assoc($result);

```

Naturalmente nel caso in cui tutti i codici ZIP siano già stati analizzati è necessario saltare la procedura di geocoding...

```
/* verifico se tutte le coordinate sono già nel database */
if (empty($vuota)) {
$msgbox= "Tutti i codici ZIP sono stati già identificati precedentemente";
} else {

$msgbox= "Coordinate recuperate con successo ed inserite nel database!";

/* esistono coordinate da recuperare */

$num=mysql_numrows($result);

echo "<b><center><font face='Arial, Helvetica, sans-serif'>Step 3: Recupero delle
coordinate geografiche</b><br>";
echo "<br><b><center><font face='Arial, Helvetica, sans-serif'>Codici ZIP
elaborati in questa sessione...</b><br>";
```

Per ogni nuovo ZIP viene utilizzata la funzione *file_get_contents* di php per recuperare la stringa contenente le informazioni cercate. Geocoder.us risponde ad un'interrogazione di questo tipo <http://geocoder.us/service/csv/geocode?zip=90210> come segue:

34.088808, -118.40612, Beverly Hills, CA, 90210

Il risultato della *file_get_contents* deve essere separato in 5 elementi con la funzione *explode* e caricato in un array (\$esploso). A questo punto è possibile effettuare una query di inserimento sulla tabella **test** e caricare i valori del nuovo codice ZIP schedato.

```
/* inizia il loop per riga */

$i=0;
while ($i < $num /*$num*/ ) {
$ZIP=mysql_result($result,$i,"FinalZIP");

/* echo "<br>$i   $ZIP<br>"; */

$coord= file_get_contents("http://geocoder.us/service/csv/geocode?zip=$ZIP");

$esploso= explode(",", $coord, 5);

/*print_r($esploso[$i]); (lat, long, city, state, geozip) */

$queryinserimento = "INSERT INTO test VALUES
('$esploso[0]','$esploso[1]','$esploso[2]','$esploso[3]','$esploso[4]')";
$resultins = mysql_query($queryinserimento);

echo "<b> $ZIP </b>";

$i++;
}
/* chiudo il ciclo else recupero zip */
}
```

Finita la fase di geocoding la pagina php visualizza come verifica il contenuto della tabella **test** aggiornata.

```
echo "
  <script type='text/javascript'>
    alert('$msgbox');
  </script>
";
/* rapido dump della tabella aggiornata per verifica */

$queryverifica="SELECT * FROM test ORDER BY geozip";
$resultver=mysql_query($queryverifica);
$numver=mysql_numrows($resultver);

if ($numver==0) {
echo "Il database sembra ancora vuoto, qualcosa non ha funzionato!";
} else {

/* inizia il loop per colonna */

$j=0;
while ($j < $numver) {

/* assegno delle variabili alle varie colonne */
$lat=mysql_result($resultver,$j,"lat");
$long=mysql_result($resultver,$j,"long");
$city=mysql_result($resultver,$j,"city");
$state=mysql_result($resultver,$j,"state");
$geozip=mysql_result($resultver,$j,"geozip");

?>
<center><table border="1" cellspacing="2" cellpadding="2"> <tr>
<th width="10%" bgcolor="#CCCCCC"><font face="Arial, Helvetica, sans-
serif">Longitude</font></th><th width="10%" bgcolor="#CCCCCC"><font face="Arial,
Helvetica, sans-serif">Latitude</font></th><th width="20%" bgcolor="#FFCC00"><font
face="Arial, Helvetica, sans-serif">City</font></th><th width="5%"
bgcolor="#CCCCCC"><font face="Arial, Helvetica, sans-serif">State</font></th>
<th width="10%" bgcolor="#0099FF"><font face="Arial, Helvetica, sans-
serif">Geozip</font></th></tr><tr><tr>
<td><font face="Arial, Helvetica, sans-serif"><? echo $lat; ?></font></td>
<td><font face="Arial, Helvetica, sans-serif"><? echo $long; ?></font></td>
<td><font face="Arial, Helvetica, sans-serif"><? echo $city; ?></font></td>
<td><font face="Arial, Helvetica, sans-serif"><? echo $state; ?></font></td>
<td><font face="Arial, Helvetica, sans-serif"><? echo $geozip; ?></font></td>
</tr>
<br>
<?
$j++;
}

echo "</table><br><br><a href=map.php>Inizia il tracciamento geografico degli
autori... <br>un marker sarà aggiunto sulla mappa per ogni
autore</a></p></center>";
/* chiudo il loop e chiudo la tabella */
}

mysql_close();
?>
```

Fase VI: mappatura geografica degli autori

Una volta ottenute le coordinate per ogni ZIP nel database è stato possibile procedere con una rappresentazione geografica degli autori. Per la cartina sono state utilizzate le API di Google Maps, facendo "scrivere" dinamicamente a php il codice Javascript necessario a visualizzare i diversi marker. La pagina in questione è <http://takeitisi.clemit.it/map.php> (per il sorgente aggiungere *?CodiceSorgente*).

Nella parte iniziale del documento si fa riferimento allo script di Google Maps, a cui viene inviata la chiave (**key**) univoca richiesta a Google per il progetto TakeitISI.

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xmlns:v="urn:schemas-microsoft-
com:vml">
<head>
  <meta http-equiv="content-type" content="text/html; charset=UTF-8"/>
  <title>Map</title>
  <style type="text/css">
v\:* {
  behavior:url(#default#VML);
}
</style>

  <script
src="http://maps.google.com/maps?file=api&v=1&key=ABQIAAAA0uG88AQtBwi2UJWYnVfaWRRW
qy5gxIzOTu-wAkG9-z12eM8EgRSKet-g8PzeAOMZq-ryjeBrSXB8jw"
type="text/javascript"></script>

</head>
<body>
  <div id="map" style="width: 980px; height: 532px"></div>
```

Nella prima query vengono collegate le due tabelle (**AU_ZIP** e **test**) sulla base del campo ZIP, in modo da poter associare ad ogni autore le coordinate di latitudine e longitudine corrette.

```
<?
if ($_SERVER['REQUEST_METHOD'] != 'POST' && isset($_GET['CodiceSorgente']))
{
  highlight_file(__FILE__);
  die();
}

include("dbdata.inc.php");
mysql_connect(localhost,$user,$password);
mysql_select_db($database) or die( "Impossibile accedere al database
TakeitISI. Verificare i parametri di connessione");

$query="SELECT * FROM AU_ZIP LEFT JOIN test ON trim(FinalZIP) = trim(geozip)
WHERE geozip IS not NULL and FinalZIP is not null GROUP BY FinalZIP ORDER BY
FinalZIP";
$result=mysql_query($query);
$num=mysql_numrows($result);

if ($num==0) {
echo "Il database sembra ancora vuoto, controllare!";
} else {

echo "<b><center><font face='Arial, Helvetica, sans-serif'><br>Step 4: Mappa degli
autori<br>click sul marker per sapere di chi si tratta</font></center></b><br>";
```

L'interprete php viene momentaneamente disattivato per inserire il codice Javascript relativo all'oggetto Map di Google Maps.

Nella prima sezione vengono aggiunti il controllo per spostarsi sulla mappa, per decidere il livello di zoom (**GSmallMapControl**) e per cambiare il tipo di mappa (**GMapTypeControl**); la finestra viene centrata sulla città di Battle Creek in Michigan ad un livello di zoom pari a 13 e viene visualizzata una finestra di benvenuto.

Successivamente vengono definite le caratteristiche dell'icona usata come marker di default (**baseIcon**) sulla mappa per indicare gli autori (immagine di sfondo, dimensioni...).

```
?>

<script type="text/javascript">
//

var map = new GMap(document.getElementById("map"));
map.addControl(new GSmallMapControl());
map.addControl(new GMapTypeControl());
map.centerAndZoom(new GPoint(-85.22294, 42.30244), 13);
map.openInfoWindow(map.getCenterLatLng(), document.createTextNode("Benvenuto
nella fase di rappresentazione grafica di TakeitISI"));

// codice di base per tutte le icone


var baseIcon = new GIcon();
baseIcon.shadow = "ombra.png";
baseIcon.iconSize = new GSize(20, 34);
baseIcon.shadowSize = new GSize(37, 34);
baseIcon.iconAnchor = new GPoint(9, 34);
baseIcon.infoWindowAnchor = new GPoint(9, 2);
baseIcon.infoShadowAnchor = new GPoint(18, 25);</pre></div><div data-bbox="113 525 888 627" data-label="Text"><p>Viene creata la funzione <b>createMarker</b>, che richiede 3 parametri: <b>point</b> (coordinate geografiche separate da virgola), <b>scrittoda</b> (autore dell'articolo, viene visualizzato quando si fa click sul marker), <b>MarkerCount</b> (numero di autori presenti in un area con uno stesso codice zip). Viene aggiunto un <b>GEvent.addListener</b> per intercettare il click del mouse sul marker ed eseguire la relativa <b>openInfoWindowsHtml</b>, che visualizza il numero degli autori nello stesso zip ed rispettivi nomi.</p></div><div data-bbox="128 651 830 881" data-label="Text"><pre>// codice per la pop-up

function createMarker(point, scrittoda, MarkerCount) {

    var icon = new GIcon(baseIcon);
    icon.image =
"http://takeitisi.clemit.it/marker.php?image=marker.png&amp;text=" + MarkerCount;
    var marker = new GMarker(point, icon);

    // visualizza autore nella pop-up
    var html = MarkerCount + " articolo/i. Autore/i &lt;b&gt;" +
        scrittoda + "&lt;/b&gt;";
    GEvent.addListener(marker, 'click', function() {
        marker.openInfoWindowHtml(html);
    });

    return marker;
}</pre></div>
```

Per poter visualizzare il numero di autori in uno stesso punto (es. ) è necessario che l'icona fondamentale di ogni marker sia generata dinamicamente da php. Quando nella prima parte del codice viene definita la **baseIcon** non è infatti presente alcun valore per la proprietà principale dell'oggetto javascript in questione: **image**. Il valore è inserito soltanto in fase di rappresentazione invocando uno script esterno.

<http://takeitisi.clemit.it/marker.php?image=marker.png&text=> + MarkerCount

```
<?php
header("Content-type: image/png");
$image_name = $_GET['image'];

$im = imagecreatefrompng($image_name);

imageAlphaBlending($im, true);
imageSaveAlpha($im, true);

$string = $_GET['text'];
//$orange = imagecolorallocate($im, 220, 210, 60);
$black = imagecolorallocate($im, 0, 0, 0);

$len = strlen($string);

if($len <= 2) {
    $px = (imagesx($im) - 7 * strlen($string)) / 2 + 1;
    imagestring($im, 3, $px, 3, $string, $black);
} else {
    $px = (imagesx($im) - 7 * strlen($string)) / 2 + 2;
    imagestring($im, 2, $px, 3, $string, $black);
}

imagepng($im);
imagedestroy($im);

?>
```

marker.php

SCRIPT realizzato da
Ali Mohammad and
Alex Gruenstein

Dopo questa prima sezione di codice Javascript il controllo torna a php, in modo da generare dinamicamente i punti sulla base delle informazioni presenti nel database. Per recuperare le informazioni necessarie a visualizzare il numero degli autori sul singolo ZIP viene effettuata una query di conteggio del valore FinalZIP sulla tabella AU_ZIP.

```
// visualizzazione marker, il controllo torna a php

<?

    $m=0;
    while ( $m < $num ) {

        $lat=mysql_result($result,$m,"lat");
        $long=mysql_result($result,$m,"long");
        $city=trim(mysql_result($result,$m,"city"));
        $state=mysql_result($result,$m,"state");
        $geozip=trim(mysql_result($result,$m,"geozip"));
        $author=mysql_result($result,$m,"AU");

        /* query sul database per recuperare il numero di autori sullo stesso marker */

        $queryMarker="SELECT COUNT(FinalZIP) as Marker FROM AU_ZIP WHERE
trim(FinalZIP)=\"$geozip\" GROUP BY FinalZIP";
        /* debug    echo ($queryMarker); */

        $resultMarker=mysql_query($queryMarker);
        $numMarker=mysql_numrows($resultMarker);
        $MarkerCount=trim(mysql_result($resultMarker,"Marker"));
```

Successivamente viene eseguita una query per recuperare tutti gli autori per un determinato codice ZIP (**\$authors**).

```
/* query sul database per recuperare tutti gli autori dello stesso marker */
$queryAU="SELECT AU FROM AU_ZIP WHERE trim(FinalZIP)=\"\$geozip\" ORDER BY AU ASC";

/* debug  echo ($queryAU); */
$resultAU=mysql_query($queryAU);
$numAU=mysql_numrows($resultAU);

    $authors = "<br>";

    for ($i=1; $i < $numAU; $i++)
    {
        $row = mysql_fetch_assoc($resultAU);
        $authors .= $row['AU'];
        $authors .= "<br>";
    }

    $row = mysql_fetch_assoc($resultAU);
    $authors .= $row['AU'];
```

Adesso che sono stati recuperati tutti i dati necessari dal database è possibile procedere con la fase finale: php scrive dinamicamente il codice Javascript per ogni marker, inserendo i valori corretti delle coordinate geografiche, degli autori e del numero di autori con lo stesso ZIP.

```
/* crea un Gpoint per ogni geozip nel database */

echo  "\n";
echo  "var point = new GPoint($long, $lat); ";
echo  "var marker = new createMarker(point, '$authors', $MarkerCount); ";
echo  "map.addOverlay(marker); ";

    $m++;
    }
}

/* chiudo il loop ed if su m */

?>

// il controllo torna a javascript

//]]>
</script>
<?
echo "<center><a href=map2.php>Inizia il tracciamento geografico dei
coautori...</a></center>";
mysql_close();
?>

</body>
</html>
```



Questo primo tipo di mappa è molto utile per verificare l'esistenza di cluster geografici dalle caratteristiche particolari o evidenziare in quali aree si concentra l'attività scientifica in un determinato settore (es. vicino alle principali università).

Fase VII: mappatura geografica dei legami tra coautori

Il legame tra coautori è di particolare interesse per verificare l'esistenza o meno di invisible collegues nell'ambito di un determinato settore. Collegando con una retta i coautori di un determinato articolo scientifico è possibile vedere immediatamente se ed in quale misura tali rapporti siano influenzati dalla prossimità geografica o la localizzazione in determinati cluster. L'ultimo file che andremo ad analizzare è <http://takeitisi.clemis.edu/map2.php> (per il sorgente aggiungere *?CodiceSorgente*). È importante notare che il richiamo al microsoft-VML (anche nel CSS) è fondamentale per rappresentare le rette correttamente in Internet Explorer.

```

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xmlns:v="urn:schemas-microsoft-
com:vml">
<head>
  <meta http-equiv="content-type" content="text/html; charset=UTF-8"/>
  <title>Map</title>
  <style type="text/css">
    v\:* {
      behavior:url(#default#VML);
    }
  </style>

  <script
src="http://maps.google.com/maps?file=api&v=1&key=ABQIAAAA0uG88AQtBwi2UJWyNVfaWRRW
qy5gxIz0Tu-wAkG9-zl2eM8EgRSKet-g8PzeAOMZq-ryjeBrSXB8jw"
type="text/javascript"></script>

</head>
<body>
  <div id="map" style="width: 980px; height: 532px"></div>

```

```

<?
if ($_SERVER['REQUEST_METHOD'] != 'POST' && isset($_GET['CodiceSorgente']))
{
    highlight_file(__FILE__);
    die();
}

/* leggo dal database i dati che mi serviranno dopo */
include("dbdata.inc.php");
mysql_connect(localhost,$user,$password);
mysql_select_db($database) or die( "Impossibile accedere al database
TakeitISI. Verificare i parametri di connessione");

    echo "<b><center><font face='Arial, Helvetica, sans-serif'><br>Step 5: Mappa
dei co-autori<br></font></center></b><br>";

?>

<script type="text/javascript">
//

var map = new GMap(document.getElementById("map"));
map.addControl(new GLargeMapControl());
map.addControl(new GMapTypeControl());
map.centerAndZoom(new GPoint(-85.22294, 42.30244), 13);
map.openInfoWindow(map.getCenterLatLng(), document.createTextNode("Ogni retta
unisce 2 coautori"));
</pre>
</div>
<div data-bbox="113 393 888 480" data-label="Text">
<p>Nella mappa dovrà essere disegnata una retta per ogni gruppo di coautori ovvero per ogni articolo. Per questo la prima query sul database va ad estrarre proprio i singoli codici <b>UTNew</b> degli articoli. Il ciclo while su <b>$pubblicazione</b> si chiuderà solo più avanti a rappresentazione dei punti conclusa. La seconda query utilizza il valore corrente di <b>$pubblicazione</b> per isolare i codici ZIP degli autori interessati e recuperarne le coordinate.</p>
</div>
<div data-bbox="126 494 869 882" data-label="Text">
<pre>
// visualizzazione marker, il controllo torna a php
&lt;?
/* iterare per singolo UT */

$queryUT="SELECT DISTINCT UTNew FROM AU_ZIP";
$resultUT=mysql_query($queryUT);
$numUT=mysql_numrows($resultUT);

$pub=0;
while ($pub &lt; $numUT) {

    $pubblicazione=mysql_result($resultUT,$pub,"UTNew");

    echo "var points$pub = []; \n";

/* iterare per ogni ZIP del singolo UT */

    $queryZIP="SELECT * FROM AU_ZIP LEFT JOIN test ON trim(FinalZIP) = trim(geozip)
WHERE geozip IS not NULL and FinalZIP is not null and
trim(UTNew)=\"$pubblicazione\"";
/* debug echo $queryZIP; */
    $resultZIP=mysql_query($queryZIP) or die( "Errore nella query");
    $numZIP=mysql_numrows($resultZIP);

    $m=0;
    while ($m &lt; $numZIP) {

        $lat=mysql_result($resultZIP,$m,"lat");
        $long=mysql_result($resultZIP,$m,"long");
        $city=mysql_result($resultZIP,$m,"city");
        $state=mysql_result($resultZIP,$m,"state");
        $geozip=mysql_result($resultZIP,$m,"geozip");
</pre>
</div>
```

Php scrive il codice Javascript dinamicamente, aggiungendo ad ogni iterazione sul codice ZIP due coordinate all'array **points\$pub** (dove **\$pub** è la variabile che segue le iterazioni su **\$pubblicazione**). Questo array sarà usato per disegnare ogni retta dalla funzione **GPolyline**.

```

echo "\n";
echo "points$pub.push(new GPoint($long, $lat)); ";

    $m++;
}
echo "map.addOverlay(new GPolyline(points$pub)); ";
echo "\n";

/* chiudo il loop su UT */
$pub++;
    }

/* chiudo il loop ed if su m - chiudo il db */
mysql_close();

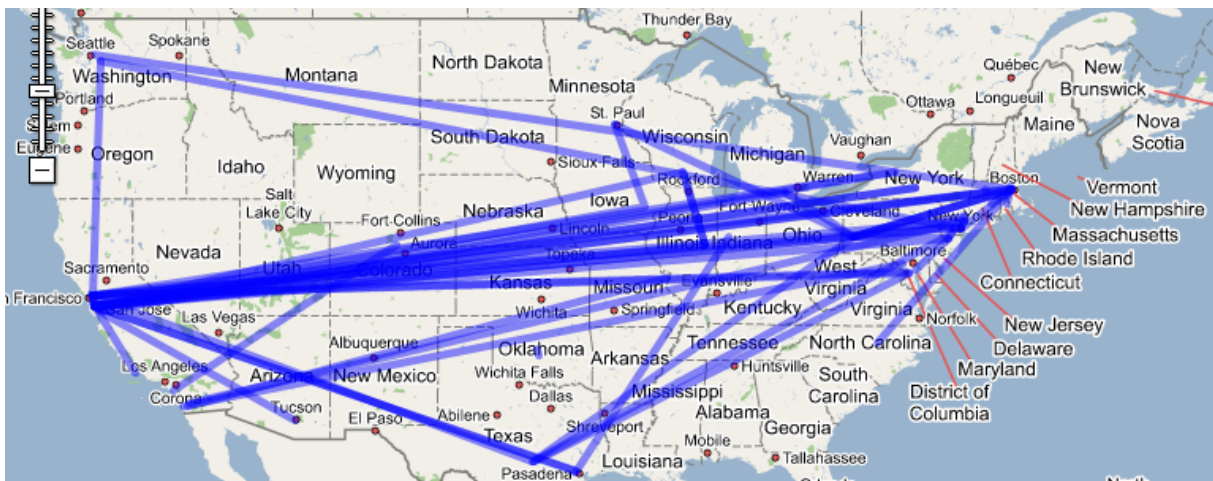
?>
// il controllo torna a javascript

//]]>
</script>

</body>
</html>

```

Il risultato finale dimostra, nel caso della rivista analizzata a scopo di esempio (Acm Transactions on Computer Systems), l'esistenza di forti legami tra le due coste degli Stati Uniti. La Silicon Valley e l'area identificata con la Route 128 confermano anche nella nostra limitata analisi il loro ruolo fondamentale nell'industria informatica.



/ fine */*