

- Unidos México información período México programa organizaciones como aplicación México programa nacionales Comisión México La General organismos no para mujer como como México sobre por la México un
9. número necesario participación principales proyectos son así informe mexicana nivel respecto Los indicadores mayor cada lo El mujer sobre como una con del se indicadores mexicana proyectos indicadores de una participación entre la a un nivel problemas mayor principales que al sobre las manera mujeres indicadores desarrollo En Presidente indicadores Se no las mexicana Unidas indicadores El sobre sus indicadores sobre proyectos Presidente las no manera necesario El indicadores proyectos indicadores no al la si indicadores indicadores por mexicana proyectos indicadores mujer desarrollo por mexicana Naciones mexicana sus deben mexicana Presidente
  10. organizaciones sistema UIT En Fuerza República dólares entre Unidos no Estados Desarrollo como Asociaciones Comité Consejo con su UNCTAD sobre Mundial que en las del los Internacional la Fuerza desarrollo UNCTAD período el Estados la Fuerza Fuerza Unidas UIT que Fuerza en Asociaciones Fuerza UNCTAD sistema dólares período El desarrollo países por Mundial con al Fuerza El el organizaciones un General UNCTAD del General Fuerza República Fuerza Naciones su UIT Fuerza no UIT Fuerza UNCTAD Se
  11. nivel respecto internacional internacionales Los lo marina sistema todos DE programas Asamblea General entre Consejo Mundial actividades países no climáticos Unidas como una por del para los en el la de climáticos marina marina R en sobre climáticos las marina Consejo los fracciones aplicación Consejo marina al período marina respecto marina entre zonas Mundial respecto marina zonas marina un sobre Gobierno climáticos todas sin sus mundial marina sus climáticos marina marina sus marina marina marina fracciones Estados del General marina por Los las En
  12. todos Los lo América celebrada recursos resolución La más aplicación octubre Carta fecha ante Secretario como sobre una Naciones cultura con al se por que el en los cultura Unidas del cultura que Naciones cultura las fecha Unidas con no fecha Naciones Secretario General octubre cultura económica fecha cultura La por todos humanos cultura Permanente con por cultura cultura lo celebrada República ante desarrollo cultura Asamblea económica cultura desarrollo países resolución Asamblea cultura por
  13. como contra han programa Bretaña Gran Iraq sesiones El Comisión proyecto Secretario Naciones sobre su E/CN desarrollo Estados por se las del el en Iraq que contra Iraq como Internacional resolución por sesiones Comisión Iraq Iraq Iraq Iraq Gran actividades Iraq por proyecto América sesiones Iraq por Bretaña Iraq Comisión Naciones Iraq por por Gran por Gran Iraq Iraq Estado más Consejo proyecto del con una Iraq Consejo que Bretaña;
  14. derechos educación internacionales sistema Lo Los años está modernas modernos nacionales parte Se aplicación este ha sus Estados así respecto también En más moderna por una un con su al del el las lo de modernas modernos otros sólo de situación modernas modernas un sólo desarrollo han no educación moderna del derechos este modernas moderna el los modernos modernos una moderna modernas lo Naciones también Lo su se El la del por modernas este Naciones Lo del moderna contra General como sistema actividades es modernos son Lo con con Lo sólo no modernos Lo su actividades Lo más Lo situación
  15. datos informe Estados nacional tradicional apoyo nacionales Los capacitación ha medios atención General mujer tradicionales países En actividades desarrollo por sobre con para del que las en los tradicionales mediante la servicios mujer datos evaluación países las tradicionales del personal por tradicionales mujer atención sobre más educación Comisión tradicionales medios tradicionales en tradicional relación su nacionales parte informe no más entre tradicionales aplicación tradicional tradicional y personal tradicionales siguientes parte con actividades tradicionales los mujer nacionales Secretario tradicionales tradicionales siguientes Comisión con zonas tradicionales atención relación Naciones sistema En tradicionales
  16. Relaciones Soviéticas Sri Tanzania Uganda Unida ha Canadá China Colombia Kuwait Noruega Repúblicas Socialistas derechos humanos Bajos agua Austria favor Consejo Japón Tratado votación India no un Conferencia Países Unido una Comisión Desarme Francia Reino Unión Arabe período Asamblea Túnez al informe El Votos Comité Irlanda proyecto Secretario resolución se contra América sesiones con por su Estados Unidos Naciones Unidas sobre México General para República que las los el en del T la r de Sri
  17. examen marco medidas nivel respecto XXXIV Programa exitosa fin informe Comité programas período han como Estados La desarrollo Naciones éxito su al un con para las de de XXXIV éxito se sobre éxito servicios exitosa XXXIV su nivel paz al éxito respecto la respecto éxito países medidas éxito éxito nacional examen nivel éxito exitosa todos éxito éxito programas al éxito nacional información respecto las su han éxito sistema sus han marco la éxito exitosa éxito han ha de programas proyecto Naciones General éxito como aeroespaciales;
  18. Conferencia Convención Asuntos sus internacionales proyecto contra países Consejo Estado sesiones Asamblea su un al República General sobre Naciones Unidas para las r de lo Asuntos Asuntos República Asuntos una Asuntos cuadragésimo países Naciones actividades República al al se han Estado Asuntos Asuntos proyecto cuadragésimo Asuntos que desarrollo Naciones Asuntos una contra el Asuntos de actividades Asuntos
  19. forma nuevas particular resultados contra económico fin lo medios Comité período sesiones Comisión internacionales Estados computadoras cuestiones mediante medio programa sistema computadora actividades El así servicios al sus países como su desarrollo una Naciones la La computadora organizaciones para actividades para computadora internacional así forma información computadora como período particular del computadoras el con Se computadoras General computadoras
- actividades programa computadora computadoras período del como computadora lo forma computadora computadora desarrollo ese servicios En computadora se sobre lo computadoras país que computadoras computadoras económico En computadora computadora computadoras
20. también todo censura cooperación cual miembros Botswana internacional Gobierno Presidente respecto lo transmitía entre han Comisión En junio Secretario una prensa se el de Botswana Botswana prensa censura informe ha al Botswana Botswana Las las dos países contra censura censura informe La Botswana Botswana ha prensa Botswana recursos Botswana sobre Estados una con para General los su Botswana que período contra Botswana programas Seguridad Botswana General cual Estados Botswana un Secretario Botswana Botswana Presidente Botswana La informe junio Botswana Secretario
  21. funcionarios otros sistema República aplicación derechos Informes La Población Social decisión lo programas todos información jefes actividades entre Desarrollo período resolución desarrollo sesiones al su General Naciones sobre los del el en la Población internacionales jefes sobre jefes con Población jefes Comité sesiones gubernamentales todos jefes al jefes jefes se jefes jefes información un un Consejo jefes lo jefes entre funcionarios Informe un lo informe jefes de el jefes de aplicación un jefes para contra programas funcionarios jefes sobre su jefes Informes Unidas Informe asistencia en funcionarios Informes una jefes sobre desarrollo resoluciones los para jefes como Programa Informes entre Población jefes desarrollo El funcionarios sus asistencia Informes organizaciones decisión informe Se otros
  22. internacionales Comité General causa Los más como actividades sobre al desarrollo una Naciones Unidas su no por del T de El causa causa sus causa que actividades causa causa causa situación información que causa no causa así las Naciones causa causa Comité causa que causa un causa como causa más como internacional causa causa las se causa La ha un causa que ;
  23. China contra cooperación ese personal parte decisión lo El internacional información representante Angola actividades proyecto años sus derechos sobre una del el la Angola conformidad Unidas el del Angola las Angola La China conformidad con China proyecto una Angola internacional Naciones China China en sobre China resolución Angola lo Angola información con los por años Angola al período información China a Angola China La respecto Angola Angola China diciembre han países Angola
  24. dólares internacionales parte particular sistema Estados gastos Comité La internacional lo es JP ha países El un al su Unidas sobre una por fuente septiembre con para del las el la de fuente dólares lo han todos fuente han fuente en Unidas país fuente por para fuente fuente Unidas fuente países dólares que fuente internacionales gastos fuente un todos medidas han fuente es fuente Estados Unidas septiembre por internacionales fuente por han fuente las Estados dólares
  25. resolución octubre medidas no como oro Carta El una DE desarrollo su países Unidos Naciones Estados las los con ante República oro oro una resolución General octubre oro resolución se América resolución oro Permanente del oro países oro un América Estados Unidas oro oro medidas oro oro los en oro En se países sobre para oro;

13. Pakistán Polonia Unión Yugoslavia como contra han miembros período programa Bangladesh Bretaña Canadá Consejo Gran Iraq Reino Unido informe sesiones Internacional sus China El India Irlanda Población actividades niños Comisión En programas países población proyecto América Unidos más no Estado Secretario General resolución Naciones Unidas sobre su una al E/CN desarrollo un con México Estados por República se para que las los del el en la de
14. Territorio acceso contabilidad derechos educación entre final hacia internacional internacionales otros qué sistema Contabilización General Lo Los años actividades comunicación contra está modernas modernos nacionales parte paz sólo técnicas Naciones Se Unidas abril aplicación después este fin ha intereses situación sus tecnologías Estados así esta han países todos La respecto son El lo también tecnología desarrollo no En sobre más como moderna por una es un con su al para se del el las que los en la de;
15. siguientes todos zonas Nairobi Oficina Se datos hacia informe jóvenes niño programas Estados deberían futuro más nacional parte personal sanitarios tradicional vivienda aplicación apoyo educación mediante nacionales no servicios sistema La Los evaluación niños primaria Comisión capacitación ha medios El México Secretario e entre relación DE Naciones Unidas información atención sus General mujer como tradicionales salud países En al una actividades un su desarrollo por sobre se con para del que las el en los la de
16. Relaciones Soviéticas Sri Tanzania Uganda Unida ha Canadá China Colombia Kuwait Noruega Repúblicas Socialistas derechos humanos Bajos agua Austria favor Consejo Japón Tratado votación India no un Conferencia Países Unido una Comisión Desarme Francia Reino Unión Arabe período Asamblea Túnez al informe El Votos Comité Irlanda proyecto Secretario resolución se contra América sesiones con Nueva por su Estados Unidos Naciones Unidas sobre México General para República que las los el en del T la r de
17. examen marco medidas nacional nivel no nuevos nutrición paz preciso respecto resultados satélite servicios tenido XXXIV apoyo espacio gestión información negociaciones posible recursos sistema Programa exitosa fin informe mediante todos Asamblea Comité Miembros aeroespaciales esfuerzos programas proyecto resolución sesiones aplicación había objetos período han internacional programa Secretario El es lo más como Consejo Estados La ha sus desarrollo países General En Naciones Unidas una éxito sobre su al por un con se para del las que los en el la de
18. Verificación ha lo note reunión India Soviética Universidad York Conferencia El actividades Convención nucleares Asuntos armas cuadragésimo programas sus Exteriores años entre han internacionales proyecto Angola Túnez contra países Relaciones humanos América derechos desarrollo Consejo Misión no Estado Secretario Unidos resolución una Nueva período Comisión Tratado Comité sesiones Asamblea por su Estados México un con al República General sobre Naciones Unidas para se que los las el del en la r de
19. ese estadísticas forma había labor nuevas país particular regionales resultados Estadística ambientales contra económico fin lo medios Comité Información Secretario internacional nacionales recursos tecnologías todos parte período programas sesiones Comisión México Se ambiente internacionales más organizaciones Estados computadoras cuestiones datos entre mediante medidas medio programa General sistema computadora actividades ha Conferencia El así no servicios En La al sus países como su desarrollo por una información Naciones Unidas un sobre con se para del que las el los en la de
20. régimen recursos también toda todo Conferencia La censura cooperación cual frontera miembros Botswana El Las Seguridad dos internacional programas CTPD Gobierno Presidente informe párrafo resolución emitido período respecto es más ha países texto lo transmitía Carta Comité comunicado entre han representante sus Comisión dirigida no sesiones En junio información marzo Estados Sudáfrica fecha Naciones Unidas contra Secretario una Consejo como sobre su prensa con General para un al se por las los del que en el la de
21. funcionarios otros sistema República aplicación derechos internacionales párrafo Informes La Población Social contra decisión mujer Administración gubernamentales lo sus Se desertificación ha informes organizaciones programas todos Internacional información jefes resoluciones El asistencia informe como actividades Comisión En entre países Secretario una Desarrollo no período resolución Comité Programa Asamblea desarrollo sesiones al un Consejo Informe su por con General Naciones Unidas se sobre para que las los del el en la de
22. así esa internacionales tierras trabajadores h Comité General causa enfermedades ha internacional todas Los lo más todos programas situación contra entre El como países La información actividades sobre sus al un desarrollo una Naciones Unidas su no por para con se del que T los el las en la de
23. votación China En contra cooperación diciembre ese nacionales personal presente conformidad pares parte período Convención La ha más Miembros Verificación decisión han lo El Misión internacional respecto Consejo Secretario información representante Angola Comisión actividades programas proyecto entre años Estado sus Asamblea República como derechos países humanos resolución no desarrollo sobre una General Estados al México Naciones Unidas su un con por para se del las los el que en la de;
24. General Se Su cuenta dólares han información internacionales país parte particular sistema todos AG Estados Gaza actividades fondos fue gastos importante medidas sido cadáver entre financiación Comité La como cooperación informe internacional lo peligrosos En México es sus JP fuentes ha países El desarrollo no un al su Unidas desechos Naciones sobre una por fuente septiembre con se para que del las el los en la de

25. of resolución Colombia Conferencia Declaración Islámica Japón LA Socialista Sudáfrica celebrada octubre La Países Soviética medidas período Alemania Italia Nueva plata India Permanente deuda En Irlanda dirigida no un Francia ante como fecha monedas oro Carta El Kuwait una Secretario DE desarrollo su América sobre México países General Unidos Naciones para Estados con se República al por las que los del el en la de

## APPENDIX IV: Queries After Optimization

1. Checoslovaquia En nacional Egipto Filipinas Portugal Finlandia gubernamentales Unidas una sesiones Mundial México resolución no un países organizaciones sus su República al sobre que en la Egipto nacional Filipinas Conferencia países México Checoslovaquia México México Egipto México México una Finlandia mujer México Egipto las se Finlandia Egipto como Comisión información E/CN sobre un Unidas General Unidas desarrollo países Finlandia Filipinas México actividades un nacional no Conferencia Filipinas Checoslovaquia Portugal nacionales Conferencia México República Egipto México al nacional proyecto México Secretario mujer que proyecto Filipinas que México Filipinas Finlandia la México En Checoslovaquia
2. Egipto Los servicios Colombia Asamblea Naciones Unidos documento su Argentina En General una al países Estados sobre un República con México del en una Colombia México servicios una México que Estados Egipto México en México siguientes Argentina trabajo Egipto México Asamblea documento Egipto Argentina República con de Secretario trabajo México principios la aplicación Colombia Argentina DE Egipto Colombia han las aplicación General Colombia Argentina servicios Colombia un documento han México los una en las México México con
3. Repúblicas Ecuador General Gran Votos sus Socialistas España Indonesia Argentina sistema organismos países China El México sobre se Irlanda Permanente Seguridad para Presidente Naciones Unidas al República por en las del de Bretaña China México México México Unidas Seguridad Argentina Gran Consejo Indonesia países Bretaña México Estados Ecuador Argentina siguiente España Gran contra Argentina Ecuador Reino Irlanda fecha Irlanda ante Votos Ecuador el México Consejo Grecia países el México Grecia Soviéticas Repúblicas México Ecuador Argentina México Socialistas Argentina sobre Ecuador el México México Bretaña Unidos sistema México Socialistas las sus
4. Nicaragua Uganda Yugoslavia Zelandia con para Botswana Finlandia Bangladesh Pakistán Repúblicas Socialistas Brasil Bretaña Gran América Argentina India Votos El Irlanda Estados México Unidos que los República de México que México Venezuela Colombia Kenya Bretaña México El Votos América Ucrania Democrática México de Pakistán Yugoslavia Botswana Soviéticas Socialistas Finlandia Socialistas Argentina Gran Argentina República Bretaña Soviéticas Nicaragua Yugoslavia Yugoslavia México Bulgaria México México Estados Bangladesh El Democrática México de Pakistán Yugoslavia Indonesia Botswana Kenya Brasil Uganda Uganda Ucrania Indonesia Socialistas Zelandia Hungría Democrática India Venezuela Repúblicas Indonesia Irlanda Venezuela Repúblicas Estados Guinea Bulgaria República Kenya Gran Democrática Irlanda México Yugoslavia Indonesia Indonesia Finlandia Bretaña del Indonesia
5. Portugal internacional República parte son Comité nacional sesiones Asamblea asistencia países período contra entre han La información como proyecto una Estados Naciones Unidas con por sobre General al un se del el en Portugal gubernamentales que Portugal nacionales República por los Portugal Portugal organizaciones la República Unidas E/CN Portugal para sesiones Conferencia situación una Portugal República DE Portugal una por los Portugal DE Conferencia un nacionales Asamblea información contra han La Portugal países Portugal sesiones Portugal su al Portugal no General contra Portugal E/CN para nacional al sobre Portugal DE Portugal un Desarrollo Desarrollo Portugal Portugal sobre por entre La un Portugal sus;
6. servicios situación tiempo Lo Qué fuerzas humanos proyecto contra más Presidente Se informe respecto Estados República Unidas derechos resolver medidas internacional lo han sobre su por con del los que el resolver tiempo derechos la humanos se Lo más Lo derechos Lo Qué Lo con Se Presidente han sobre fin al una resolver Lo La Naciones medidas resolver Lo ha por sus Unidas tiempo resolver Lo todos Presidente organizaciones Naciones resolver mujeres internacional proyecto fuerzas organizaciones resolver internacional una fuerzas al la nacionales Asamblea Lo internacional respecto plazo Qué Comité han Consejo por Asamblea derechos Lo con las Qué El Lo como Lo Unidas El con Unidas Se han Qué las han El como Lo General fuerzas plazo resolver con
7. han Comisión todos información presente miembros es actividades sesiones armas Conferencia Secretario Naciones Unidas con por sobre al para las que el la de Comité Unidas armas armas armas del armas los armas armas presente período todos armas E/CN en armas armas que Estados armas Mundial organizaciones se han han programa armas como diciembre para Unidos período armas armas miembros un informe armas es General armas Conferencia diciembre;
8. América La gubernamentales medidas como Unidos información presente México Mundial aplicación desarrollo actividades Asamblea sistema un Conferencia organizaciones General al sobre del que las el la México Mundial México América República período Unidos México por México México Unidos la México su República período una Asamblea México su México México México informe por período actividades México actividades Unidos la General República sesiones del

## APPENDIX I: Spanish Queries

1. Indicaciones de las relaciones económicas y comerciales de México con los países europeos.
2. Indicaciones de las relaciones económicas y comerciales de México con los países africanos.
3. Indicaciones de las relaciones económicas y comerciales de México con los países asiáticos, por ejemplo Japón, China y Corea.
4. Indicaciones de las relaciones económicas y comerciales entre México y Canadá.
5. ¿Hay programas e intercambios deportivos entre México y los Estados Unidos?
6. ¿Cuáles son las medidas tomadas por el gobierno mexicano para resolver la disputa con los rebeldes zapatistas en el estado de Chiapas?
7. ¿Cuál es la importancia de las Naciones Unidas (NU) para México?
8. ¿Cómo van las relaciones entre México y la Organización de los Estados Americanos (OEA)?
9. Indicaciones de los potenciales y debilidades del ejército mexicano.
10. Indicaciones de los potenciales y debilidades de las fuerzas aéreas militares de México.
11. Indicaciones de los potenciales y debilidades de la marina de guerra (fuerzas navales, armada) de México.
12. Evidencia de la herencia y cultura azteca en México.
13. ¿Hay programas para la renovación urbana en México?
14. ¿Cuáles son las medidas modernas para mejorar la agricultura en México?
15. Información sobre el ballet folklórico en México.
16. Medidas para controlar o evitar inundaciones en México.
17. ¿Tendrá éxito el NAFTA (TLC) en México?
18. Hay programas para reprimir o limitar epidemias en México?
19. Información sobre la industria de computadoras mexicana.
20. Actitudes en México sobre la censura de la prensa.
21. Informes sobre visitas oficiales y privadas a México por jefes de estado y de gobierno.
22. Hay programas en México para investigar la causa de cáncer?
23. Hay programas internacionales para el intercambio de estudiantes en México?
24. El turismo como fuente de divisas para México.
25. La fabricación en México de joyas de plata y oro.

## APPENDIX II: English Queries

1. Indicators of economic and business relations between Mexico and European countries.
2. Indicators of economic and business relations between Mexico and African countries.
3. Indicators of economic and business relations between Mexico and Asian countries, such as Japan, China and Korea.
4. Indicators of economic and business relations between Mexico and Canada.
5. Are there sports programs and exchanges between Mexico and the United States?
6. What measures has the Mexican government taken to resolve the quarrel with the rebel Zapatistas in the state of Chiapas?
7. What importance does the United Nations (UN) have for Mexico?
8. How are relations between Mexico and the Organization of American States (OAS)?
9. Indicators of the Mexican Army's strengths and weaknesses.
10. Indicators of the Mexican Air Force's strengths and weaknesses.
11. Indicators of the Mexican Navy's (naval and marine forces) strengths and weaknesses.
12. Evidence of Aztec heritage and culture in Mexico.
13. Are there urban renewal programs in Mexico?
14. What modern measures for agricultural improvement are there in Mexico?
15. Information about Mexico's traditional dance (ballet folklorico).
16. Flood prevention and control measures in Mexico.
17. Will NAFTA (TLC) be successful in Mexico?
18. Are there epidemic control programs in Mexico?
19. Information about the Mexico's computer industry.
20. Attitudes in Mexico regarding censorship of the press?
21. Reports regarding official and private visits to Mexico by chiefs of state and heads of government?
22. Does Mexico have research programs for the cause of cancer?
23. Are there international student exchange programs in Mexico?
24. Tourism as a source of Mexico's income?
25. Silver and gold jewelry manufacturing in Mexico?

## APPENDIX III: Queries Before Optimization

1. Checoslovaquia En Ghana Polonia nacional programa Australia Bajos Egipto España Filipinas La Países Portugal Igualdad Italia Paz recursos Austria Finlandia Acción Pide Venezuela Naciones gubernamentales Unidas como período una Comisión Desarrollo regionales sesiones Mujer Mundial información nacionales informe México resolución no proyecto un actividades países Estados organizaciones desarrollo sus su E/CN mujer Secretario General por República al con se Conferencia sobre para del las que los el en la de
2. Checoslovaquia Democrática Egipto Filipinas Francia Indonesia Irlanda Los Países Secretario Uruguay aplicación más proyectos servicios Alemania Colombia La fuentes trabajo Asamblea Iraq Naciones Nigeria Pakistán Unidos documento han DE Unidas energía nuclear sus Brasil principios siguientes utilización Argentina Chile En Venezuela como desarrollo espacio ultraterrestre El General una período sesiones al países su Estados sobre un para República por con se México que las del los en el la de
3. Italia Repúblicas Chile Ecuador General Gran India Pakistán Suecia Unión Venezuela Votos sus Bretaña Filipinas Norte Nueva Socialistas Soviéticas siguiente Alemania Bajos Colombia España Grecia Indonesia Reino su Argentina Francia Japón Unido especializados Brasil Países sistema Arabe organismos países Nicaragua Sudáfrica China organizaciones contra El México América sobre se Representante ante Irlanda con Permanente Unidos Seguridad para Estados Carta fecha dirigida Presidente Consejo Naciones Unidas al que República por en las los del el la de
4. Nicaragua Nigeria Uganda Yugoslavia Zelandia con favor Dinamarca Indonesia Norte Ucrania Venezuela Yemen para Australia Botswana Finlandia Italia Soviéticas Suecia Austria Bangladesh Bulgaria Chile Hungría Irán Islámica Noruega Pakistán Repúblicas Socialistas Bajos Brasil Bretaña España Gran Países Reino Unión Unido Egipto Francia Perú Polonia América Argentina India Japón Kenya Canadá Checoslovaquia China Alemania Guinea Nueva Colombia Democrática Votos El Irlanda Arabe Estados México Unidos que las los el del en la República de
5. Misión Pide Portugal Sudáfrica Verificación aplicación internacional programa situación todos Angola Desarrollo República gubernamentales parte son Comité Consejo Mujer diciembre ha nacional regionales sesiones En Asamblea DE asistencia Estado Miembros países período programas contra entre han nacionales La Comisión desarrollo información sus como organizaciones actividades informe proyecto una Conferencia mujer Estados Naciones Unidas su Secretario no resolución E/CN con por sobre General al un se para del las los el que en la de
6. saber servicios situación tiempo Comisión Consejo Departamento Israel Lo Qué Unidos agua fuerzas funciones humanas mujeres nacionales organizaciones proyecto Asamblea México plazo pueblo En Gobierno contra desarrollo más otros refugiados todos Presidente Se como fin informe parte respecto Comité Estados General Naciones República Unidas derechos resolver La medidas internacional lo sus han países sobre El ha su no es una un al por para con se del las los en que el la de
7. entre han Comisión Comité Convención Mundial Paz Unidos organizaciones todos Desarme Desarrollo OOPS información presente América miembros República medidas La mujer regionales como es Asamblea actividades una programa proyecto desarme período diciembre sesiones armas informe sus Conferencia nucleares E/CN Secretario resolución no un Naciones Unidas su con por General sobre Estados al se para del las los que en el la de
8. nacionales programa regional América La gubernamentales medidas En como Unidos información desarme especializados presente Comisión México mujer armas Pide informe organismos Mundial aplicación desarrollo nucleares diciembre una Desarrollo actividades sesiones Asamblea período regionales sistema países resolución República Secretario no sus un Conferencia organizaciones Estados se su General por con al Naciones Unidas sobre para del que las los el en la de
9. medida metas número necesario participación período principales programas proyectos pueden puntos recursos regional regionales respecta si son tener ventajas vez Se así biotecnología deben determinar financiera informe manera mexicana nivel positivos respecto sistema situación Los gestión indicadores mayor mecanismos país Naciones República actividades cuenta deficiencias entre examen más mujeres Unidas cada organizaciones problemas La Presidente es lo En El al México mujer países sobre como desarrollo no sus un una con su por del se para el que las los en la de
10. Preparatorio Presidente Se actividades Educación General Los Oficina organizaciones sistema OACI UIT Conferencia En Fuerza Movimiento República desarrollo países período Internacional contra dólares entre Ciencias Europea Unidos no Estados OMI Desarrollo El como Asociaciones Sociedad Comité Naciones Unidas una sus al un OIT FAO Unión Consejo Instituto ONUDI con Organización su UNCTAD sobre se por Mundial Asociación OMS que UNESCO Federación en las del para los el y Internacional la de
11. sin Meteorológica examen nivel nota posible respecto todas En deben es internacional internacionales mundial zonas Ambiente Gobierno Los Medio aplicación lo más marina período sistema Organización fracciones recursos también todos Administradora La DE Potencia así programas Asamblea General entre medidas Territorio Consejo Mundial Programa organizaciones desarrollo El actividades países Estados no al climáticos Naciones cambios Unidas como sus su una un sobre se por con del para que las los en el la de
12. todos económica transmite Colombia España Los Sudáfrica contra lo medidas pueblos América Representante Venezuela celebrada recursos resolución Asamblea Declaración La coloniales más República aplicación humanos México deuda es internacional octubre Carta Permanente derechos fecha dirigida territorio un Estados no ante desarrollo sus Secretario como sobre países Unidas una Naciones cultura General para su con al se por las del que el en los la de

*Annual Conference on Evolutionary Programming*, San Diego, Evolutionary Programming Society, 1995.

Dunning, T. E., and M. W. Davis (1993a), A Single Language Evaluation of a Multilingual Text Retrieval System. In NIST Special Publication 500-207: *The First Text Retrieval Conference (TREC-1)*, ed. D.K. Harman, Computer Systems Laboratory, NIST.

Dunning, T. E., and M. W. Davis (1993b), Multilingual Information Retrieval. *Memoranda in Computer and Cognitive Science*, MCCS-93-252, Computing Research Laboratory, New Mexico State University.

Fogel, D. B. (1992), A Brief History of Simulated Evolution. In *Proc. of the First Annual Conference on Evolutionary Programming*, ed. D.B. Fogel and J.W. Atmar, 1-16. San Diego: Evolutionary Programming Society.

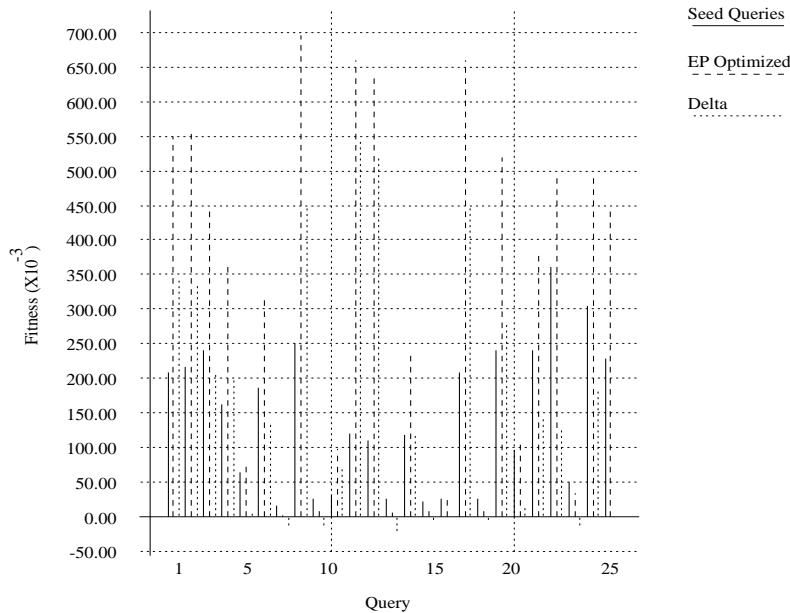
Harman, D. K. (1993). Overview of the First Text Retrieval Conference. In NIST Special Publication 500-207: *The First Text Retrieval Conference (TREC-1)*, ed. D.K. Harman, Computer Systems Laboratory, NIST.

Landauer, T. K. and M. L. Littman (1990). Fully Automatic Cross-Language Document Retrieval Using Latent Semantic Indexing. In *Proceedings of the 6th Conference of UW Centre for the New Oxford English Dictionary and Text Research*, 31-38. Waterloo.

Salton, G. (1970). Automatic Processing of Foreign Language Documents. *Journal of the American Society for Information Sciences*, 21: 187-194.

Salton, G. and M. J. McGill (1983), *Introduction to Modern Information Retrieval*. New York: McGraw Hill.

### Performance Improvements in EP-MLTR



**Figure 4:** Performance of seed queries, EP-optimized queries and difference between them over the novel Spanish corpus.

On average, the dictionary-based queries produced performance which was about 50% worse than the reference queries. The EP-derived queries produced performance which was 60-70% worse than the reference queries. Surprisingly, in some cases, performance of the dictionary-based or EP-derived queries was better than the original Spanish queries. The reasons for this anomaly are not completely understood; they could be due to statistical fluctuation, or it could be that the translation process performed useful query expansion in these cases.

## 9.0 Conclusions and Future Work

This work demonstrates that a relatively simple evolutionary programming strategy that eliminates and replaces terms in queries can produce machine translated queries for use in IR systems. Much work remains to be done, however, and evaluation methods must still be improved. As with any EP methodology, the approach presented in this paper was resource intensive. The evaluations were drawn from an extremely large text corpus of 1.5 gigabytes of documents. Optimization on a fast Sparcstation 20 with the corpus on local disk required several hours. As expected, this somewhat limits the viability of this approach for on-line IR systems. For standard routing and document filtering tasks where queries are updated relatively infrequently (or in the context of a personal information filtering agent that runs over-

night), the delays associated with translation and optimization may be acceptable.

Finally, the extremely large processing delays may be mediated by the use of standardized query engines with better performance.

## 10.0 Acknowledgments

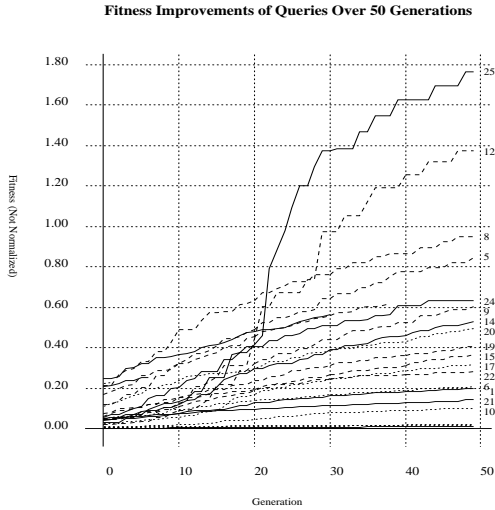
This research was funded under the U.S. Department of Defense Tipster development effort, Contract No. MDA 904-94-C-6153-3.

## 11.0 References

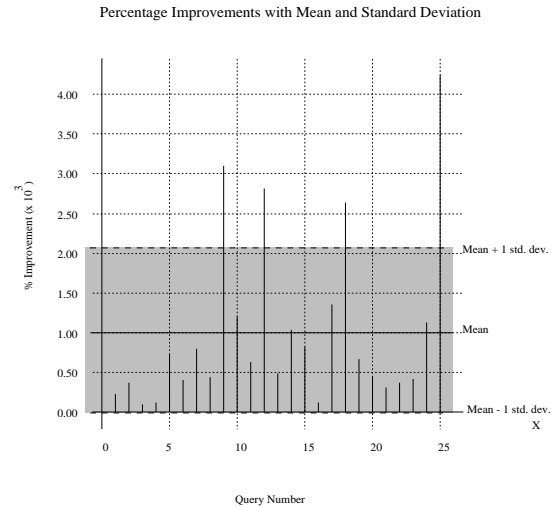
Davis, M. W. and T. E. Dunning (1995a), A TREC Evaluation of Query Translation Methods for Multilingual Text Retrieval. In NIST Special Publication: *The Fourth Text Retrieval Conference (TREC-4)*, ed. D.K. Harman, Computer Systems Laboratory, NIST.

Davis, M. W., T. E. Dunning, and W. C. Ogden (1995) Text Alignment in the Real World: Improving Alignments of Noisy Translations Using Common Lexical Features, String Matching Strategies and N-Gram Comparisons. In *Proceedings of the Conference of the European Chapter of the Association of Computational Linguistics*. University College Dublin. March 1995.

Davis, M. W. and T.E. Dunning (1995b) Query Translation Using Evolutionary Programming for Multilingual Information Retrieval. In *Proceedings of the Fourth*



**Figure 2:** Fitness improvements of queries 1-25 over 50 generations during training.



**Figure 3:** Percentage Improvements for Queries 1 through 26, with mean and standard deviation indicated, due to evolutionary programming (training phase).

as a baseline for comparing other retrieval results. A further assumption that is made is that a hand-generated translation of a query will retrieve identical documents on an English portion of the corpus as the original Spanish form. If the English query does not, then optimizing queries against the results from the English side of the training corpus should not necessarily result in any improvement of the query when compared to the original Spanish query on the novel document set.

The improvements from the training phase do appear to result in improvements over novel corpora. Figure 4 shows the fitness values (normalized) of the seed query with respect to original Spanish query over the novel corpus, as well as the improved query fitness, and the difference between the seed and improved query fitness values.

Queries 7, 9, 13, 15, 16, 18 and 23 all showed either decreased performance or no improvement due to optimization. Except for queries 9 and 15, both of which showed moderate gains during optimization, the remaining queries all had little or no improvement during the optimization phase. Query 7 is, at first glance, something of an oddity, since the query concerns the United Nations. The presence of “United Nations” in the query, however, appears to have had little impact on optimization improvements. This can likely be explained by the fact that “United Nations” occurs more or less uniformly throughout the corpus and the existence of the phrase in a query therefore would have little effect.

## 8.0 Evaluation on TREC documents

In addition to using held out portions of the UN corpus for evaluation, the EP-derived query translations were compared to the original Spanish queries by participating in the comparison of Spanish information retrieval systems which was part of the fourth TREC. The other systems in this comparison were all Spanish monolingual systems, although many were still called multilingual.

The CRL submissions to TREC-4 included a number of different approaches to query translation which were described in Davis and Dunning (1995a). The discussion here is limited to comparison of a reference baseline (original Spanish queries), the dictionary-based query translations, and the EP-derived query translations. The reference set of queries were the original TREC-Spanish queries in their original form. The dictionary-based queries were generated by using a machine readable dictionary to produce all possible term translations for terms in the English versions of the Spanish queries. The EP-derived queries were the translations of the English versions of the queries described earlier in this report.

Each query in each of these three sets was evaluated according to the TREC guidelines as *ad hoc* queries. The Inquiry retrieval engine by University of Massachusetts, Amherst, was used to perform the retrievals, and all results were scored using the TREC pooled evaluation methodology which is described in (Harman, 1993).

potential alignment block matches. The approach used in this work supplements the approach in Davis, Dunning and Ogden (1995) by automatically identifying alignment “anchors” based on character 4-grams prior to performing the maximum-likelihood test over the regions between the anchor points. Tests of 500 randomly chosen alignment pairs demonstrated an 83% alignment match-rate over text that had not been pre-processed, and was known to contain massive formatting irregularities.

## 6.0 Implementation

Twenty-five Spanish Queries were evaluated. The Spanish queries (Appendix I) are standard queries that were created by National Institute of Standards (NIST) for the evaluation of Spanish IR systems in the Text Retrieval Evaluation Conference (TREC). The queries were translated by a Spanish-literate graduate student to make standard English versions of the queries (Appendix II).

For simplicity, the original Spanish queries will be referred to as OS and the hand-translated English versions as OE in the descriptions that follow. The training corpus, consisting of around 80,000 alignment segments will be called TC with the English component noted as TC<sub>E</sub> and the Spanish component as TC<sub>S</sub>. The evaluation corpus will be called EC, with EC<sub>E</sub> and EC<sub>S</sub> components, respectively.

The seed query set (Appendix III) was generated by running the English queries against TC<sub>E</sub> to create a ranked document-query vector. The Spanish parallel texts from TC<sub>S</sub> for the top 10% of the documents retrieved were then filtered to remove high-frequency terms and the remaining terms were lumped together, sorted by frequency and the top 100 terms became the new seed query. Some formatting information was subsequently filtered from the query set (Appendix III).

The seed queries were then run individually against EC<sub>S</sub> to produce a document-query vector, as were the terms in OS.

Evolutionary optimization began with the seed query set, with each query being run against TC<sub>S</sub>, and the fitness being calculated by a dot product of the English query over TC<sub>E</sub> and the Spanish query over TC<sub>S</sub>. A population of 50 copies of each seed query formed the initial population and mutation operations involved deleting terms from the query, or adding terms from the original seed query. This dramatically limited the range of variation possible in the modified query to encompass only the terms expressed by the top 100 filtered query terms. The number of mutations per query-generation was allowed to randomly vary between 1 and 10 addition/deletions per query-gener-

ation, although more sophisticated strategies are certainly possible.

The variation in fitness, measured by the *unnormalized* document-query vector dot product, is shown in Figure 2. The variation of fitness improvements is substantial for many queries. Furthermore, it appears that fitness improvements could be expected beyond 50 generations since most of the queries that show large improvements do not appear to have peaked. The mean improvement over all 25 queries was 1000% with a substantial standard deviation of 1043.8%, as shown in Figure 3.

An analysis of the effort applied by the evolutionary methods is informative. The seed queries in Appendix III have a mean length of 76 terms, resulting in:

$$\sum_{i=1}^{76} \binom{76}{i} \cong 7.5 \times 10^{22}$$

deletion operations possible for the average seed query. Under the further assumption that an average of five mutations were applied to each query each generation, that means that the evolutionary programming method explored only 2,500 queries over 50 generations, or  $0.33087 \times 10^{-17}$  % of the total search space for each query.

## 7.0 Evaluation Over Novel Corpora

In order to evaluate the performance of the new queries generated by the evolutionary programming methods, each improved query was run against the remaining 600,000 documents of the corpus that had been excluded from the training phase. The original Spanish queries were also run against this portion of the corpus, as were the seed queries for the evolutionary methods. The similarity of the improved query-document vector and the original Spanish query, as compared to the similarity of seed query-document vector to the original Spanish query yields a measure of the improvements by the evolutionary programming optimization method.

This is not the most ideal method for evaluation. In most IR evaluation processes, ranked retrieval results (query-document vectors) are evaluated by human judges to obtain a precision-recall curve (Salton and McGill, 1983). The precision measures the judges’ collective judgment as to the relevancy of the retrieved documents to the query. The recall measures the judges’ opinions as to the coverage of the retrieved documents over all relevant documents.

If it is assumed that a human-generated query produces reasonable results in a fairly well-evaluated system like the IDF-weighting system used in the results presented herein, then those results can be considered

shows simple vector spaces for documents constructed out of three terms over three documents.

In an evolutionary programming (EP) (Fogel, 1992) context, the comparative evaluation of document score vectors serves as an objective measure of the relative fitness of a query to the collection:

$$f = \Delta_i \bullet \Delta_j$$

For the EP approach, an initial population of queries is needed along with a mutation strategy to modify queries. Optimization then proceeds by evaluating the comparative fitness of the queries, mutating a selected sub-population of the queries to produce “offspring” solutions and re-evaluating the queries iteratively until a suitable number of generations have passed.

#### 4.0 Deriving Initial Queries Via Parallel Corpora

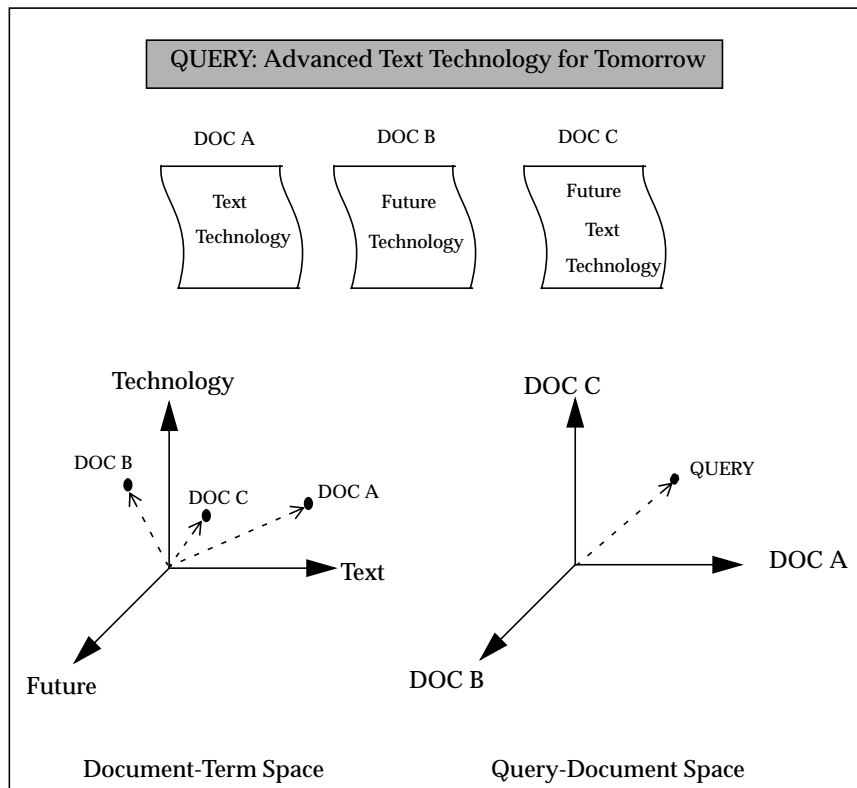
In contrast to the method reported in Davis and Dunning (1995b), in this work the approach to creating initial queries used terminology drawn directly from a parallel training corpus. Whereas machine readable dictionaries produce translations that are narrow and

comprehensive, a parallel corpus produces text translations that are wide and shallow, consisting of domain-specific usage patterns combined with additional general terminology that may not be directly related to the terms of interest. Document retrieval results for each English query were obtained and the parallel Spanish text for the top ten English documents were sorted by frequency after being filtered to remove high-frequency terms (the top 250 terms from the training corpus). The method is described in some detail below.

#### 5.0 Parallel Corpora and Alignment

The availability of parallel, aligned corpora is critical to corpus-based translation efforts. For this work, a collection of 683,000 aligned text segments were automatically compiled from 4 years of United Nations reports. The reports detailed UN proceedings, sponsored initiatives and other day-to-day activities of the UN and member countries. The domain of the documents was therefore predominantly international development and global policy issues.

The alignment algorithm is, in part, described in Davis, Dunning and Ogden (1995) and, in brief, applies a maximum-likelihood statistical test to select among



**Figure 1:** Document-term and Query-Document vector space representations for three-documents containing three-terms.

system used a thesaurus for generating query translations by taking the terms in the thesaurus for each query term and forming a new translated query. The thesaurus was created by hand for the retrieval corpus and the entries were therefore inherently disambiguated with respect to the corpus domain prior to query generation. Nevertheless, Salton's results demonstrate that IR systems can perform well in a multilingual setting using simple translation resources. Unfortunately, domain-specific, up-to-date glossaries are generally difficult to obtain. Those that are produced are typically constructed by and for translators, who write them in the process of translation, suggesting that an approach which makes use of the translations directly in combination with other resources is needed.

Experiments with latent semantic indexing (LSI) (Landauer and Littman, 1990) showed that paragraphs which were translations of each other could be correlated but no actual retrieval system was constructed, nor was it clear how the system would perform in practice. This use of parallel corpora eliminates many of the problems of using bilingual dictionaries, but introduces new problems. In particular, in the context of a traditional vector based retrieval system, it has not been clear how to perform multilingual retrieval based on the information contained in parallel translated corpora. The success of experiments with LSI does not directly provide a method to make a more traditional vector based system work. Furthermore, LSI makes the use of inverted indexes problematic, which may hinder the practicality of this system.

Dunning and Davis (1993a,b) developed a system for multilingual information retrieval based on a novel method for solving very large systems of linear equations. In this system, query translation was viewed as a linear transformation of a query feature vector. For long strings, the translation of the concatenation of the strings is approximately the translation of the strings independently. This is true because the translation of two strings is nearly the concatenation of their translations. While this linearity breaks down dramatically at the word level, at the sentence level and above, it works fairly well. Despite the simplification afforded by linearity in the transformation, the actual translation matrix was derived through a computationally taxing error minimization strategy that used a parallel aligned corpus of 50,000 words as exemplars to iteratively update the transformation matrix. At that time, machine resources were very limited and the algorithm had poor convergence properties.

Davis and Dunning (1995b) applied an evolutionary approach to parallel collections by optimizing for translated query performance over a collection of parallel texts. The initial queries were created from term-by-term lookup in a bilingual machine readable dictio-

nary. Mutation operations included adding or deleting terms from the pooled sense terminology for the dictionary entry. Although promising, the work left many unanswered questions about the usefulness of general lexicons for highly specific text domains and the value of corpus-based term disambiguation in an IR framework. Moreover, the results appeared to be affected by the lack of high-quality parallel text for training and evaluation.

### 3.0 A Corpus-based Evolutionary Approach to Query Translation

In the vector IR model of information retrieval, both queries and documents are represented by vectors of weights derived from term occurrence patterns in the texts. Given a text,  $T$ , a vector can be constructed

$$t(T) = [w_1 \ w_2 \ \dots \ w_n]$$

where the weights are derived from a statistical measure of term occurrence in the text. A classical measure is the Inverse Document Frequency (IDF) measure (Salton, 1988), scaled by term frequency:

$$w_{ij} = f_{ij} \log\left(\frac{N}{n_j}\right)$$

where  $N$  is the number of documents in the corpus,  $n_k$  is the number of documents that contain term  $j$  and  $f_{ij}$  is the frequency of term  $j$  in the  $i$ th document.

For document and query vectors constructed over the same basis of terms, a measure of similarity,  $d_{nq}$ , of the document and query can be determined by an inner product of their vectors. The cosine dot product is regularly used:

$$d_{nq} = t_n \cdot t_q = \frac{\sum w_{ni} w_{qi}}{\|t_n\| \|t_q\|}$$

A document score vector,  $\Delta_q$ , for the query,  $t_q$ , is then created by concatenating the results of the similarity measures over the entire document set:

$$\Delta_q = [d_{1q} \ d_{2q} \ \dots \ d_{nq}]$$

One measure of the similarity between two queries is then an inner product (typically the cosine dot product) of the document score vectors for the two queries.

This approach regards the documents and queries as existing in a space of term weights,  $S = \mathfrak{R}^m$ , where  $m$  is the number of unique terms in the corpus. Individual queries also exist in a space delineated by their similarities to documents in the collection,  $D = \mathfrak{R}^n$ , where  $n$  is the number of collection documents. Figure 1

# Query Translation Using Evolutionary Programming for Multilingual Information Retrieval II

**Mark W. Davis**

Computing Research Laboratory  
New Mexico State University  
Box 30001/3CRL  
Las Cruces, NM 88003-0001

**Ted E. Dunning**

Computing Research Laboratory  
New Mexico State University  
Box 30001/3CRL  
Las Cruces, NM 88003-0001

## Abstract

Multilingual information retrieval (IR) systems apply queries in one language to a document collection in several different languages with the goal of retrieving only those documents relevant to the query. At first glance, deep linguistic analysis and translation of the query appears necessary before retrievals can be performed. IR systems are unique in natural language processing, however, because a pattern of term occurrences in a document generally suffices to determine the subject matter; word order is largely irrelevant. Translated queries are therefore primarily derived by a mapping from a word set in the query language to a word set in the language of the derived query. This paper follows upon the work reported in Davis and Dunning (1995b). Evolutionary Programming is again applied to optimize the performance of translated queries, but the initial queries are derived from simple corpus statistics, rather than from a prepared lexicon. Results indicate substantial performance gains can be made in many cases by a simple strategy of eliminating and repeating terms from the queries.

## 1.0 Introduction

The goal of information retrieval (IR) is to retrieve documents that are closely related to a user's needs. For performance and simplicity, most systems avoid sophisticated linguistic analysis of the documents.

Some systems go so far as to consider documents to be unordered bags of words. These documents are related to a query which is itself composed of unordered words to which numeric weights are attached to indicate importance. Systems which primarily combine these weights from the query with word counts from the documents in a linear fashion are called vector-

based retrieval systems. The relevancy of a document to a query is determined by some sort of normalized inner product on the document and query vectors. Until recently, such systems represented the state of the art in retrieval for moderate to large sets of documents. Even now, vector systems that use short phrases in addition to single words are as effective as any others. (Harman, 1993).

In general, the problem of translating the user's needs into queries for such systems is central to the field of information retrieval (Salton, 1970). The query may contain many more words than appear in the original user-input. In addition, users are generally rather poor at assigning weights to the terms in the query. Choosing good terms and weighting them is hard enough in one language, but when the problem is extended to involve documents in multiple languages it becomes considerably more difficult. In one possible scenario, the user generates input for the system in only a single language, but expects to retrieve documents in multiple languages (multilingual information retrieval or MLIR). In the context of conventional vector-based retrieval systems, this could be accomplished by automatically translating all of the documents into a single language when indices of the documents are created, or by translating the user-input into a multilingual query.

This paper describes the results of work on an approach to MLIR that builds multilingual queries using a corpus of previously translated documents. The fitness of a query is determined by the comparative performance of the original query and potential translation over parallel training documents. Incremental improvements in a query's performance are propagated into subsequent generations in a simulated evolution system.

## 2.0 A Brief History of Multilingual Information Retrieval

Salton (1970) first demonstrated that information retrieval could be used in a multilingual setting. His